

UNIwersYTET EKONOMICZNY W POZNANIU
WYDZIAŁ INFORMATYKI I GOSPODARKI ELEKTRONICZNEJ
KATEDRA INFORMATYKI EKONOMICZNEJ

ROZPRAWA DOKTORSKA

AUTOMATYCZNE PRZETWARZANIE RECENZJI KONSUMENCKICH
DLA OCENY UŻYTECZNOŚCI PRODUKTÓW I USŁUG

Jacek Małyszko

Promotor: prof. dr hab. Witold Abramowicz, prof. zw. UEP

Promotor pomocniczy: dr Agata Filipowska

POZNAŃ 2015

Spis treści

1	Wprowadzenie	1
1.1	Motywacja	1
1.2	Cel badawczy i teza pracy	7
1.3	Metody badawcze i źródła wykorzystywanych informacji	10
1.4	Struktura rozprawy	14
2	Recenzje konsumenckie w Internecie i ich rola w podejmowaniu decyzji przez uczestników wymiany rynkowej	16
2.1	Geneza zjawiska	17
2.2	Fenomen zaangażowania konsumentów w publikowanie recenzji	18
2.3	Portale udostępniające możliwość publikowania recenzji	19
2.4	Formaty wyrażania opinii	23
2.4.1	Wypowiedź tekstowa jako sposób wyrażenia opinii	30
2.5	Recenzje konsumenckie a zachowania podmiotów na rynku	30
2.5.1	Wpływ recenzji na decyzje zakupowe konsumentów	30
2.5.2	Reakcje sprzedawców na fenomen recenzji konsumenckich dostępnych w Sieci	33
2.5.3	Problem przesytu informacyjnego w kontekście wykorzystania recenzji konsumenckich przez podmioty uczestniczące w wymianie rynkowej	40
2.6	Podsumowanie	40
3	Analiza wydźwięku i metody konstruowania leksykonów sformułowań wyrażających opinie	43
3.1	Drażnienie opinii	44
3.2	Określanie aspektów recenzowanych przedmiotów	44
3.2.1	Aspekt produktu - przegląd definicji	44

3.2.2	Leksykony nazw aspektów	47
3.3	Podejścia do analizy wydźwięku	48
3.3.1	Statystyczna klasyfikacja tekstów i nadzorowane uczenie maszynowe	48
3.3.2	Analiza wydźwięku z wykorzystaniem reguł i leksykonów sformułowań służących do wyrażania opinii	54
3.4	Automatyczne pozyskiwanie leksykonów sformułowań służących do wyrażania opinii	55
3.4.1	Problem względności polarności sformułowań	56
3.4.2	Identyfikacja sformułowań służących do wyrażania opinii	58
3.4.3	Identyfikacja polarności – metody słownikowe	59
3.4.4	Identyfikacja polarności – metody wykorzystujące analizę korpusu	61
3.5	Drażnienie opinii w tekstach w języku polskim	65
3.5.1	Różnice pomiędzy językami a metody i zasoby analizy wydźwięku	65
3.5.2	Opublikowane prace z zakresu analizy wydźwięku w polskich tekstach	66
3.5.3	Rozwiązania komercyjne dotyczące analizy wydźwięku dla języka polskiego	68
3.6	Zidentyfikowane braki w istniejących metodach	69
3.7	Podsumowanie	70
4	Wykorzystanie podsumowań opinii	
	o produktach i usługach do identyfikacji polarności – zarys podejścia	71
4.1	Nadzorowanie odległe a identyfikacja polarności – problem z założeniami	72
4.2	Zarys proponowanej metody	73
4.3	Warianty proponowanej metody	74
4.3.1	Wariant I	74
4.3.2	Wariant II	77
4.3.3	Wariant III	82
4.4	Założenia metody	84
4.5	Podsumowanie	85
5	Analiza spójności różnych sposobów wyrażania opinii o produktach i usługach	86
5.1	Motywacja badań	86
5.2	Koncepcja eksperymentów	89
5.2.1	Cele eksperymentów	89
5.2.2	Posiadane dane i procedura ich przygotowania	90

5.3	Eksperyment 1. Spójność informacji na poziomie pojedynczych recenzji	92
5.3.1	Koncepcja eksperymentu	93
5.3.2	Procedura eksperymentu i uzyskane wyniki	94
5.3.3	Interpretacja wyników	95
5.4	Eksperyment 2. Badanie spójności opinii wielu recenzentów	97
5.4.1	Kryterium spójności postrzegania produktów przez recenzentów	99
5.4.2	Miary odległości między podsumowaniami	100
5.4.3	Procedura eksperymentu i uzyskane wyniki	101
5.4.4	Interpretacja uzyskanych wyników	106
5.5	Podobieństwo podsumowań opinii wyrażanych w różnych formatach	106
5.5.1	Podsumowania recenzji tekstowych a podsumowanie list zalet i wad	107
5.5.2	Koncepcja eksperymentów	108
5.5.3	Procedura eksperymentów i uzyskane wyniki	108
5.5.4	Interpretacja wyników	111
5.6	Podsumowanie	112

6 Ekstrakcja sformułowań służących do wyrażania opinii i wykorzystanie częściowo strukturyzowanych recenzji konsumenckich dla identyfikacji ich polarności 113

6.1	Opis rozwiązania	113
6.2	Wymagania dla metody	115
6.3	Przygotowanie leksykonu aspektów i ich nazw	117
6.3.1	Identyfikacja aspektów	117
6.3.2	Leksykon nazw aspektów	118
6.4	Ekstrakcja sformułowań służących do wyrażania opinii	119
6.4.1	Analiza korpusu	120
6.4.2	Przygotowane rozwiązanie	120
6.4.3	Ewaluacja rozwiązania	126
6.5	Identyfikacja polarności	129
6.5.1	Wejście i wyjście dla metody	129
6.5.2	Przypisywanie polarności	130
6.6	Podsumowanie	131

7	Ewaluacja wypracowanej metody	132
7.1	Eksperyment 1. Przypisywanie polarności na podstawie podsumowań list zalet i wad	133
7.1.1	Dane	133
7.1.2	Wypracowana metoda	135
7.1.3	Metody bazowe	135
7.1.4	Procedura przeprowadzenia eksperymentu	138
7.1.5	Uzyskane wyniki i ich interpretacja	139
7.2	Eksperyment 2. Wykorzystanie skonstruowanego leksykonu w analizie wydźwięku	144
7.2.1	Przebieg eksperymentu i wykorzystane dane	145
7.2.2	Otrzymane wyniki i ich interpretacja	146
7.3	Eksperyment 3. Identyfikacja polarności na podstawie podsumowań ocen punktowych	147
7.3.1	Identyfikacja restauracji	147
7.3.2	Przebieg eksperymentu i wykorzystane dane	148
7.3.3	Uzyskane wyniki i ich interpretacja	151
7.4	Podsumowanie	152
8	Podsumowanie	156
8.1	Wkład pracy	156
8.2	Dalsze badania	159
	Bibliografia	162
	Załączniki	177
A	Adnotacja opinii wyrażanych przez recenzentów w dokumentach tekstowych – wskazówki adnotacji	177
A.1	Cel procesu adnotacji	177
A.2	Typy adnotacji	177
A.3	Wskazówki adnotacji	177
B	Budowanie leksykonu nazw aspektów na podstawie ręcznej analizy korpusu – wskazówki adnotacji	179
B.1	Cel procesu adnotacji	179
B.2	Aspekty i typy adnotacji	180
B.3	Wskazówki adnotacji	182

C	Budowanie bazy przykładowych sformułowań używanych do oceny aspektów produktów – wskazówki adnotacji	185
C.1	Cel procesu adnotacji	185
C.2	Typy adnotacji i powiązania z aspektami produktów	185
C.3	Wskazówki adnotacji	187
D	Oznaczanie poprawności wyekstrahowanych fraz używanych do oceny aspektów produktów – wskazówki adnotacji	189
D.1	Cel procesu adnotacji	189
D.2	Wskazówki adnotacji	190
D.3	Przykładowe adnotacje	191
E	Polarność sformułowań służących do wyrażania opinii – wskazówki adnotacji . . .	194
E.1	Cel procesu adnotacji	194
E.2	Dane do adnotacji	194
E.3	Wskazówki adnotacji	195
F	Analiza wydźwięku – wskazówki adnotacji	198
F.1	Cel procesu adnotacji	198
F.2	Dane do adnotacji	198
F.3	Wskazówki adnotacji	198
G	Przykład wykorzystanej reguły ekstrakcji w formalizmie JAPE	201

Spis rysunków

1.1	Macierz obrazująca różne rodzaje wkładu do nauki w ramach badań z informatyki ekonomicznej	8
1.2	Cykle w procesie badań nad systemami informatycznymi z zakresu projektowania	11
2.1	Fragment pierwszego poziomu drzewa kategorii produktów na portalu cokupic.pl	20
2.2	Podsumowanie wielu ocen punktowych w postaci rozkładu liczby recenzentów, którzy przypisali produktowi określoną liczbę punktów	25
2.3	Fragment podsumowania recenzji gry komputerowej The Sims 2 na portalu cokupic.pl	26
2.4	Materiał marketingowy ze strony brand24	38
3.1	Schemat klasyfikacji dokumentów tekstowych z wykorzystaniem nadzorowanego uczenia maszynowego	50
4.1	Proces wyliczania statystyki umożliwiającej przeprowadzenie identyfikacji polarności na podstawie podsumowań wielu opinii o recenzowanym bycie dla pierwszego wariantu proponowanej metody	75
4.2	Zrzut ekranu z portalu gastronauci.pl prezentujący przykład podsumowania opinii recenzentów o różnych aspektach wybranej restauracji	76
4.3	Przykład recenzji wyrażonej jednocześnie w postaci opisu tekstowego i listy zalet i wad	77
4.4	Proces generowania cech w proponowanej metodzie (wariant II)	82
4.5	Zrzut ekranu z systemu OpinionObserver, obrazujący podsumowania opinii wielu recenzentów o różnych aspektach trzech ocenianych telefonów komórkowych . . .	83
5.1	Rozkłady wszystkich ocen punktowych nadanych różnym kategoriom produktów .	88

5.2	Zrzut ekranu z portalu cokupic.pl prezentujący fragmenty listy zalet i wad, za pomocą której recenzenci mogą wyrażać swoje opinie o produktach z kategorii „Telefony komórkowe”	91
5.3	Wartości statystyki U w testach istotności dla dwóch średnich przy różnych populacjach produktów o niewielkiej liczbie recenzji	103
5.4	Skumulowane rozkłady odległości pomiędzy podsumowaniami tych samych produktów oraz pomiędzy podsumowaniami produktów różnych	105
5.5	Skumulowane rozkłady odległości pomiędzy podsumowaniami tych samych produktów oraz pomiędzy podsumowaniami produktów różnych, gdzie podsumowania zostały wygenerowane na podstawie różnych sposobów wyrażania opinii – wypowiedzi tekstowej oraz listy zalet i wad	110
6.1	Schemat opracowanego rozwiązania	114
7.1	Proces ekstrakcji leksykonu i przygotowania danych do identyfikacji polarności w pracy Jürgena Broña, skąd zaczerpnięto metodę wykorzystaną w eksperymentach jako jedną z metod bazowych	136
7.2	Rozkład liczby par (<i>sformułowanie, aspekt</i>) o liczbie wystąpień nie większej niż dana wartość w rozbiciu na pary przypisane przez adnotatorów do różnych klas	144
7.3	Proces przygotowywania danych dla potrzeb eksperymentu, za pomocą którego symulowano sytuację opisaną w opisie pierwszego wariantu proponowanej metody	149
7.4	Precyzja i pełność proponowanej metody dla polarności pozytywnej i negatywnej w zależności od przyjętej wartości oceny granicznej i w zależności od minimalnej liczby wystąpień par poddawanych klasyfikacji	155

Spis tabel

2.1	Porównanie różnych formatów wyrażania opinii	29
2.2	Sytuacje, w których podmioty uczestniczące w wymianie rynkowej mogą zetknąć się ze zjawiskiem przesytu informacyjnego w związku z analizą recenzji konsumenci- kich	42
4.1	Podsumowanie przyznanych zalet i wad (na listach zalet i wad) dla wybranych aspektów telefonu komórkowego Alcatel-ot311	79
5.1	Statystyki korpusu zaadnotowanych recenzji	92
5.2	Statystyki zgodności list zalet i wad z adnotacjami tekstowymi pochodzącymi ze złotych standardów	94
5.3	Fragmety podsumowań wygenerowanych na podstawie dwóch rozłącznych zbiorów recenzji telefonu komórkowego Samsung U700	98
5.4	Tablica odległości dla wszystkich par z iloczynu kartezjańskiego zbiorów podsumowań $X \times Y$	99
5.5	Porównanie liczby recenzji, w których wymienione zostały wybrane aspekty produktu bądź w opisie tekstowym, bądź w liście zalet i wad	107
6.1	Zidentyfikowane sekwencje słów, nazw aspektów i części mowy o największym pokryciu dla analizowanego korpusu, dla których przygotowano reguły ekstrakcji . .	122
6.2	Ewaluacja wyników ekstrakcji fraz służących do wyrażania opinii	128
7.1	Najczęściej ekstrahowane pary (<i>sformułowanie, aspekt</i>) w przyjętym korpusie recenzji	134
7.2	Statystyki precyzji, pełności i miary F dla poszczególnych polarności dla trzech porównywanych metod bez uwzględnienia częstości występowania poszczególnych par w korpusie	140

7.3	Statystyki precyzji, pełności i miary F dla poszczególnych polarności dla trzech porównywanych metod z uwzględnieniem częstości występowania poszczególnych par w korpusie	141
7.4	Liczebności par o danej liczbie wystąpień w analizowanym korpusie, dla których metody bazowe nie posiadały przesłanek do przypisywania polarności do par (<i>sformułowanie, aspekt</i>)	142
7.5	Liczba niezgodnych klasyfikacji między metodą bazową 1 a metodą proponowaną w rozprawie w zależności od liczby wystąpień par	143
7.6	Statystyki precyzji, pełności i miary F dla poszczególnych polarności, obrazujące poprawność analizy wydźwięku przeprowadzonej z wykorzystaniem leksykonów będących wynikiem trzech porównywanych metod	154
1	Lista aspektów produktów (telefonów komórkowych), dla których w wyniku procesu adnotacji wygenerowane mają zostać słowniki nazw	181
2	Lista aspektów produktów (telefonów komórkowych) wraz z odpowiadającymi im typami adnotacji oznaczającymi pozytywne lub negatywne postrzeganie danego aspektu przez recenzenta	186

Rozdział 1

Wprowadzenie

1.1 Motywacja

Recenzje konsumenckie publikowane w Internecie pełnią coraz istotniejszą rolę w procesie podejmowania decyzji przez uczestników wymiany rynkowej [Zhu i Zhang 2010]. Przez recenzje konsumenckie rozumiemy opinie na temat dóbr (podobnie jak w pracy [Rekowski 2008], tym zbiorczym terminem określamy produkty i usługi), opublikowane w pewnym źródle internetowym, wyrażone przez podmioty niebędące ekspertami w danej dziedzinie [Mudambi i Schuff 2010]. Autorami takich opinii często są konsumenci, czyli członkowie gospodarstw domowych, dążący do zaspokojenia potrzeb osobistych lub potrzeb innych członków gospodarstwa domowego. Opinie w postaci recenzji mogą być wyrażane również przez inne podmioty¹, na przykład nabywców niebędących konsumentami².

Recenzje konsumenckie wykorzystywane są przez podmioty będące uczestnikami wymiany rynkowej w celu optymalizacji prowadzonych przez nie działań. Sposób wykorzystania recenzji przez takie podmioty jest różny w zależności od charakteru, w jakim uczestniczą one w rynku. Inną funkcję w podejmowaniu decyzji recenzje pełnią dla kupujących (nabywców), a inną dla sprzedających. Istnieją również podmioty, które w stosunku do danego produktu lub usługi występują na rynku w obu tych rolach (pośrednicy). Również wśród podmiotów pełniących te same role, sposób wykorzystania recenzji może być różny, w zależności od przyjętej strategii działania

¹Pomimo tego, że opisywane recenzje mogą być przygotowywane nie tylko przez konsumentów, w literaturze najczęściej stosowana jest nazwa „recenzje konsumenckie”; w prezentowanej pracy również będziemy korzystać z tego określenia, ze świadomością dokonywania pewnego uproszczenia.

²Podmioty takie mogą wykorzystywać zakupione produkty i usługi przykładowo w procesie produkcji lub w celu świadczenia usług

i celu prowadzonej działalności.

Z punktu widzenia nabywców, recenzje konsumenckie mogą pozwolić im na racjonalizację decyzji zakupowych. Nabywcy będący pośrednikami mogą kupować produkty i usługi w celu dalszego ich odsprzedania. Inni nabywcy mogą również kupować produkty i usługi w celu produkcji kolejnych produktów lub świadczenia usług. W obu tych przypadkach, zapoznanie się z opiniami o kupowanych dobrach ma umożliwić wybór takich dóbr, który pozwoli im na maksymalizację osiągniętych korzyści.

Nabywcy mogą kupować dobra również w celu ich konsumpcji. Mogą oni wówczas wykorzystać recenzje konsumenckie dla dokładniejszego oszacowania tego, jaką użyteczność mogą im przynieść poszczególne produkty i usługi [Zhu i Zhang 2010]. Użyteczność wyraża subiektywny stosunek konsumenta do oferowanych na rynku koszyków produktów i usług [Panek 2000]. Jest ona „sumą zadowolenia, jaką osiąga indywidualny konsument z konsumowania lub posiadania danego dobra” [Rekowski 2008]. Każdy konsument ma swoje indywidualne preferencje, które warunkują, jaką użyteczność przynoszą mu określone dobra. Podejmując wybory, konsumenci dążą do maksymalizacji użyteczności [Rekowski 2008]. Często, informacje na temat analizowanych produktów i usług pełnią kluczową rolę w procesie oceny użyteczności jeszcze przed podjęciem decyzji zakupowej. Jest to o tyle istotne, że o ile informacja dotycząca różnic w cenach produktów i usług może być stosunkowo łatwo pozyskana, znacznie trudniejsze może być poznanie informacji pozwalających na oszacowanie, jaką użyteczność mogą te produkty i usługi przynieść konsumentom [Shapiro 1982; Smallwood i Conlisk 1979]. Bez możliwości oszacowania użyteczności, którą konsumentowi może przynieść dane dobro, nie jest możliwe właściwe ustosunkowanie się przez niego do ceny, za jaką dobro to jest oferowane na rynku. Recenzje konsumenckie, na drodze wymiany informacji o produktach i usługach pomiędzy nabywcami, mogą pozwolić na przezwyciężenie takich trudności [Hu, Liu i Zhang 2008]. Liczne opublikowane wyniki badań (przykładowo w pracach [Gretzel i Yoo 2008; The Nielsen Company 2009]) wskazują na wysoką wagę, jaką z tego powodu recenzjom konsumenckim przypisują konsumenci.

Z drugiej strony, coraz istotniejszą rolę recenzje konsumenckie pełnią w podejmowaniu decyzji przez podmioty znajdujące się po drugiej stronie rynku w stosunku do nabywców, tj. przez producentów, sprzedawców oraz usługodawców produktów i usług. Analizując recenzje, podmioty te mogą poznać opinie nabywców o produktach i usługach swoich lub konkurentów [Chen i Xie 2005; Dellarocas, Zhang i Awad 2007; Godes i in. 2005]. Opinie takie, publikowane w postaci recenzji, można traktować jako wyraz satysfakcji nabywców z recenzowanego produktu lub usługi. Satys-

fakcję rozumiemy tu jako percepcję danej osoby dotyczącą stopnia, w jakim jej wymagania zostały przez dany produkt lub daną usługę spełnione [Niezurawski, Pawłowska i Witkowska 2010]. Dzięki monitorowaniu publikowanych w Internecie recenzji konsumenckich dotyczących danego przedsiębiorstwa lub produktów i usług, w handlu którymi dane przedsiębiorstwo pośredniczy lub które wytwarza bądź świadczy, może ono zyskać informacje pozwalające na wprowadzanie modyfikacji swojego zachowania (np. odpowiednio zmienić cenę czy dostosować działania marketingowe do sposobu postrzegania dobra na rynku) oraz być w stanie lepiej dbać o swój wizerunek w Internecie, przykładowo odpowiadając na pojawiającą się w recenzjach konsumenckich krytykę.

Analizując wpływ recenzji konsumenckich na decyzje podejmowane przez podmioty będące uczestnikami wymiany rynkowej należy zwrócić uwagę na fakt, że zapoznanie się ze wszystkimi recenzjami danego produktu czy usługi opublikowanymi w Internecie może być zadaniem bardzo czasochłonnym lub nawet praktycznie niewykonalnym. Istnieją wyspecjalizowane portale, w których recenzowane są tysiące produktów lub usług, a poszczególne produkty lub usługi mogą być oceniane w wielu recenzjach. Przykładowo, na portalu gastronauci.pl w dniu 10 czerwca 2014 było ponad 47000 recenzji dotyczących 832 restauracji działających na terenie Warszawy, a niektóre restauracje miały ponad 2000 recenzji. Z punktu widzenia pojedynczego, potencjalnego klienta restauracji, liczba taka czyni irracjonalnym zapoznawanie się ze wszystkimi opiniami opublikowanymi na wspomnianym portalu.

Oczywiście, nie zawsze konieczne jest zapoznanie się z dużą liczbą recenzji. Przykładowo, w opisaney powyżej sytuacji, potencjalny klient restauracji może zdecydować się na przeprowadzenie ograniczonej analizy, w trakcie której zapozna się tylko z próbką recenzji ograniczonej listy restauracji i na tej podstawie dokona wyboru lokalu, do którego się uda. Dodatkowo, wyspecjalizowane portale służące do publikowania recenzji często pozwalają recenzentom na wyrażanie opinii w łatwym do automatycznego przetworzenia formacie, na przykład oceny punktowej (często stosowaną skalą jest liczba punktów od 1 do 5, gdzie 1 to ocena najgorsza, a 5 to ocena najwyższa). Takie oceny punktowe mogą być łatwo podsumowywane, przykładowo poprzez wyliczenie średniej spośród wszystkich ocen nadanych danemu produktowi czy usłudze. Podsumowania takie następnie mogą być publikowane na wspomnianych portalach, dzięki czemu możliwe jest szybkie poznanie przez użytkowników średniej oceny nadanej poszczególnym produktom i usługom. W niektórych przypadkach, dla jednego produktu generowanych może być wiele podsumowań, gdzie każde podsumowanie odpowiada pewnej specyficznej grupie recenzentów (na przykład, w recenzjach hoteli takie grupy mogą odpowiadać zadeklarowanym przez poszczególnych recen-

zentów typom podróży, np. podróż służbowa, wypoczynek itd.). Dzięki takim podsumowaniom, użytkownik chcący wybrać restaurację może wybrać przykładowo ten lokal, który ma najwyższą średnią ocenę.

Są jednak sytuacje, w których wskazane jest zapoznanie się z jak największą liczbą recenzji, a dostępne są wyłącznie recenzje w postaci trudnych do automatycznego przetworzenia wypowiedzi tekstowych. Taka sytuacja ma miejsce przykładowo wówczas, gdy pewne przedsiębiorstwo monitoruje określony serwis społecznościowy w celu identyfikacji wypowiedzi, w których użytkownicy negatywnie odnoszą się do jego produktów, usług czy marki. Celem monitorowania może być przykładowo umożliwienie odpowiedniemu działowi firmy reagowania na wspomniane negatywne wypowiedzi, na przykład poprzez publikowanie odpowiedzi, w których firma broniłaby się przed kierowanymi pod jej adresem zarzutami³. W takich sytuacjach, kluczowa jest szybka identyfikacja jak największej liczby wypowiedzi zawierających negatywne opinie na dany temat. Biorąc pod uwagę fakt, że na wielu portalach społecznościowych liczba wypowiedzi publikowanych każdego dnia może być liczona w milionach (np. na portalu Twitter średnia liczba wypowiedzi każdego dnia wynosi około 500 milionów⁴), ich monitorowanie przez człowieka, czy nawet duży zespół ludzi, staje się niemożliwe. Dodatkowo, poniesione koszty, wynikające z poświęcenia czasu na wyszukiwanie i analizę recenzji konsumenckich, mogą okazać się wyższe od przewidywanych korzyści, jakie mogą być dzięki temu osiągnięte.

Opisana sytuacja jest przykładem paradoksu przesytu informacyjnego. Występuje on wówczas, gdy ilość informacji na dany temat przekracza możliwości jednostki w zakresie jej przetworzenia. Decyzje podejmowane w sytuacji przesytu informacji mogą nawet okazać się gorsze, niż w przypadku jej niedoboru [Ho i Tang 2001]. Problem przesytu informacyjnego w przypadku recenzji konsumenckich może zostać łatwo rozwiązany wówczas, gdy dostępne są opinie wyrażone w łatwo przetwarzalnej postaci, np. w postaci ocen punktowych, która to postać umożliwia łatwe wygenerowanie podsumowań. Jednakże, jak opisano w poprzednim akapicie, takie łatwo przetwarzalne opinie nie zawsze są dostępne – często mamy do czynienia wyłącznie z opiniami wyrażonymi w postaci wypowiedzi tekstowych. W związku z tym, coraz istotniejsze staje się automatyczne przetwarzanie recenzji konsumenckich, przykładowo w celu generowania ich podsumowań czy szybkiego wykrywania negatywnych wypowiedzi.

Obszar badań zajmujący się opisaną problematyką nazywany jest drażeniem opinii (ang. *opi-*

³Innym celem, dla którego firma może monitorować opinie, jest identyfikowanie pozytywnie ocenianych produktów i usług, przykładowo w celu odpowiedniego ich wyeksponowania.

⁴<https://blog.twitter.com/2013/new-tweets-per-second-record-and-how>. Dostęp 27: kwietnia 2014

nion mining) lub analizą wydzźwięku (ang. *sentiment analysis*) i jest dobrze znanym problemem z zakresu przetwarzania języka naturalnego. Z formalnego punktu widzenia, opinię można tu zdefiniować jako piątkę uporządkowaną [Liu 2011]:

$$(e_i, a_{ij}, o_{ijkl}, h_k, t_l) \tag{1.1}$$

gdzie:

- e_i jest bytem, którego dotyczy opinia, tj. produktem, usługą, osobą [Silva, Carvalho i Sarmiento 2012], zdarzeniem, organizacją czy określonym zagadnieniem [Ding, Liu i Yu 2008], np. sytuacją polityczną w kraju [Mullen i Malouf 2006]; w prezentowanej pracy zazwyczaj będziemy się skupiać na dwóch rodzajach bytów: produktach i usługach, które nazywać będziemy mianem dóbr,
- a_{ij} jest pewnym aspektem bytu e_i , gdzie przez aspekt bytu będziemy rozumieć wszystko, co może wpływać na opinię innych podmiotów o tym bycie i co jest z nim bezpośrednio powiązane; na przykład, dla laptopa aspektem mogą być rozdzielczość wbudowanego monitora czy długość życia jego baterii [Liu 2011];
- o_{ijkl} to tzw. polarność opinii sformułowanej w stosunku do aspektu a_{ij} bytu e_i , przyjmująca jedną z trzech wartości: pozytywną, negatywną lub neutralną, opcjonalnie z różnymi poziomami intensywności (inne nazwy tego elementu piątki to orientacja semantyczna, orientacja wydzźwięku czy orientacja opinii) [Hatzivassiloglou i McKeown 1997; Turney i Littman 2003]; stanowi ona wyraz użyteczności, jaką byt e_i w aspekcie a_{ij} przyniósł podmiotowi wyrażającemu opinię,
- h_k oznacza podmiot (osobę lub organizację), który wyraża opinię [Ding, Liu i Yu 2008],
- t_l reprezentuje moment, w którym opinia została wygłoszona [Liu 2011].

Z punktu widzenia drażenia opinii, każdą recenzję konsumencką można zaprezentować jako kolekcję takich właśnie piątek uporządkowanych.

Często, oprócz wyrażania opinii o poszczególnych aspektach bytu, wyrażana jest też opinia o całym bycie, odwzorowująca ogólne wrażenia podmiotu o ocenianym bycie. Takie ogólne wrażenia o ocenianym bycie można traktować jako jego kolejny aspekt [Liu i in. 2014], co pozwala łatwo opisać tę sytuację.

Dla dokładniejszego wyjaśnienia poszczególnych elementów omawianej piątki uporządkowanej, warto jest przeanalizować przykładowe zdanie i określić, w jaki sposób można je zaprezentować z jej pomocą. Przykładem może być tu następujące zdanie, które mogłoby pojawić się w recenzji konsumenckiej dotyczącej pewnej restauracji:

Obsługa w lokalu była przesympatyczna, ale jedzenie niestety było dość mdłe.

Na przykładzie tego zdania, poszczególne elementy piątki uporządkowanej przedstawionej w równaniu 1.1, można rozumieć następująco ⁵:

- e_i (byt, którego dotyczy opinia) – ocenianym bytem jest pewna restauracja, w recenzji której znalazła się przytoczona wypowiedź,
- a_{ij} (aspekt bytu e_i) – ocenie podlegają tu dwa aspekty – *obsługa lokalu* oraz *jedzenie*,
- o_{ijkl} (polarność opinii sformułowanej w stosunku do aspektu a_{ij}) – w stosunku do wymienionych aspektów, mamy do czynienia z przeciwnymi polarnościami; do aspektu *obsługa* przypisana jest pozytywna polarność (wyraża ją sformułowanie *przesympatyczna*), natomiast do aspektu *jedzenie* przypisano polarność negatywną (której nośnikiem jest sformułowanie *dość mdłe*),
- h_k (podmiot wyrażający opinię) – podmiotem wyrażającym opinię jest tutaj autor recenzji. W niektórych przypadkach, na przykład w artykułach prasowych, przytaczane są opinie osób innych niż autor artykułu; wówczas automatyczne ustalenie tego elementu piątki staje się znacznie trudniejszym zadaniem niż w recenzjach konsumenckich,
- t_l (moment, w którym opinia została wygłoszona) – godzina i data opublikowania opinii (ponownie, w tekstach innych niż recenzje konsumenckie ten element piątki może być znacznie bardziej skomplikowany do ustalenia).

Bardzo istotną rolę w analizie przytoczonego zdania odgrywały sformułowania *przesympatyczna* oraz *dość mdłe*. To właśnie te sformułowania⁶, które będziemy w dalszej części określać

⁵Należy zaznaczyć, że poniższy sposób reprezentacji opinii w ramach piątki uporządkowanej z równania 1.1 jest tylko jednym z wielu możliwych. Każdy system, w zależności od postrzegania danej domeny, może przygotować tutaj różne reprezentacje tej samej opinii.

⁶Pod zbiorczym terminem *sformułowania* będziemy rozumieli zarówno pojedyncze słowa, jak i wielowyrazowe frazy.

mianem **sformułowań służących do wyrażania opinii**, pozwoliły na określenie polarności opinii dla aspektów zawartych w tekście. Aby poprawnie ocenić polarność opinii, konieczne jest posiadanie przez system wiedzy dotyczącej tego, jaką polarność w stosunku do analizowanych aspektów mogą nieść te sformułowania.

Zależnie od podejścia obranego podczas przeprowadzania analizy wydziwisku, wymagane jest wykorzystanie różnych zasobów i technik w celu rozpoznawania opinii wyrażanych w tekście. Takie zasoby to m.in. zaadnotowane korpusy czy leksykony sformułowań używanych przez recenzentów do wyrażania opinii [Liu 2011].

Prezentowana praca dotyczy zagadnienia analizy wydziwisku. W świetle przedstawionej motywacji, problemem badawczym, który praca usiłuje rozwiązać, jest zmniejszenie przesyłu informacyjnego związanego z nadmiarem recenzji konsumenckich publikowanych w Internecie, uniemożliwiającego podmiotom będącym uczestnikami wymiany rynkowej właściwe przetworzenie tych recenzji i optymalizację decyzji podejmowanych na ich podstawie.

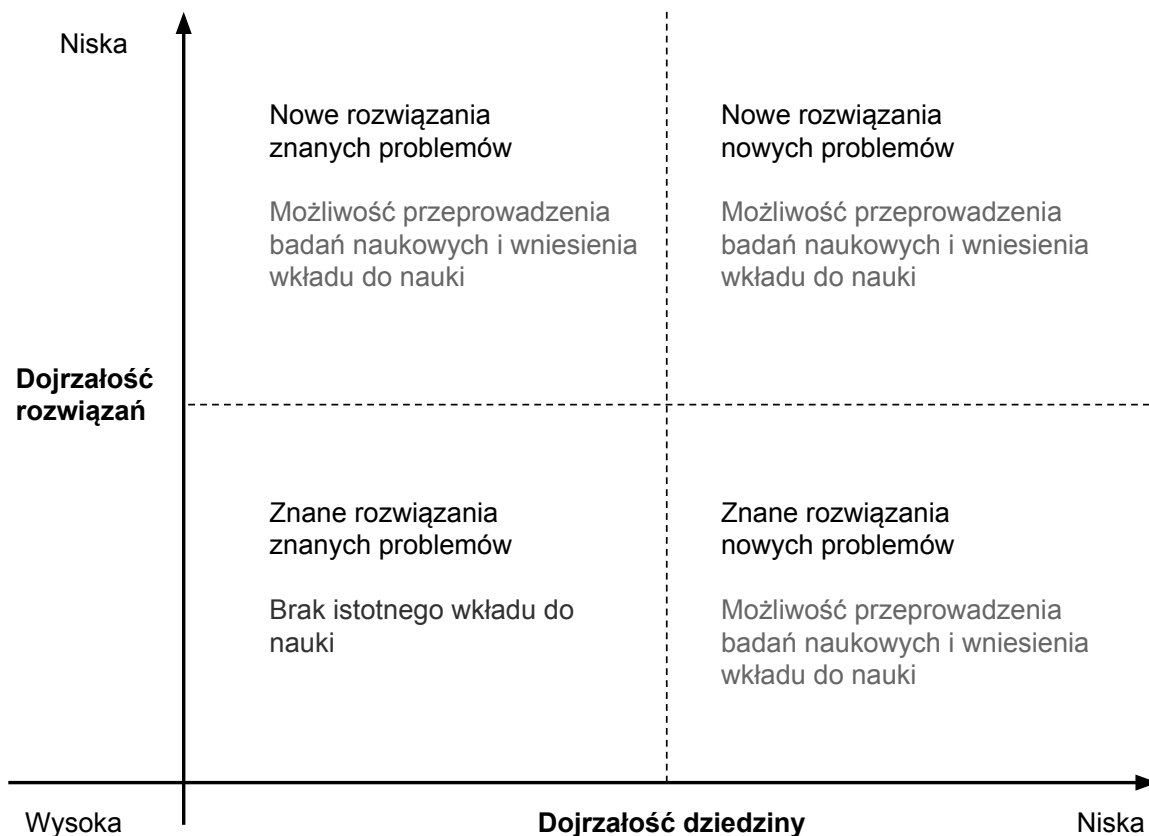
1.2 Cel badawczy i teza pracy

Celem rozprawy jest **opracowanie metody, wykorzystującej informacje zawarte w wielu częściowo strukturyzowanych recenzjach konsumenckich, w celu poprawy skuteczności istniejących metod identyfikacji polarności sformułowań służących do wyrażania opinii, dla potrzeb oceny użyteczności produktów i usług**. Przez polarność rozumiemy jedną z trzech wartości: pozytywną, negatywną lub neutralną, odpowiadającą opinii, jaką dane sformułowanie wyraża w zdaniu w kontekście danego bytu i aspektu⁷. Znajomość polarności sformułowań służących do wyrażania opinii może zostać wykorzystana do określania opinii recenzentów o produktach, usługach i różnych ich aspektach.

Prawidłowe określenie polarności sformułowań służących do wyrażania opinii jest nietrywialnym wyzwaniem. Jest to spowodowane tym, że sformułowania takie zmieniają polarność w zależności od tego, jaki byt oraz jaki aspekt opisują. Przykładowo, sformułowanie „nieprzewidywalny” zmienia swoją polarność z pozytywnej w przypadku opinii o fabule filmu („nieprzewidywalna fa-

⁷Istnieją prace, gdzie polarność rozumie się w kategoriach bardziej rozmytych, gdy pomiędzy wspomnianymi trzema wartościami może się znajdować wiele stanów pośrednich. Założenie takie może być bliższe rzeczywistości, jednak w przekonaniu autora pracy może utrudnić przeprowadzenie obiektywnej ewaluacji wypracowanych metod ze względu na większą dyskusyjność poprawności przypisywanych wartości. W związku z tym, w prezentowanej pracy polarność rozumiana jest jako posiadająca jedną z trzech wartości: pozytywna, negatywna lub neutralna.

buła”⁸ na negatywną w przypadku wyrażania opinii o samochodzie i łatwości kierowania nim („nieprzewidywalne zachowanie samochodu na zakręcie”)⁹ [Liu 2011; Lu i in. 2011]. Proces nauki przez system tego, jaką polarność mają różne sformułowania używane w różnych kontekstach, będziemy nazywać **identyfikacją polarności** lub **nauką polarności**.



Rysunek 1.1: Macierz obrazująca różne rodzaje wkładu do nauki w ramach badań z informatyki ekonomicznej. Źródło: [Gregor i Hevner 2013]

Jednym z podejść do identyfikacji polarności sformułowań jest nadzorowanie odległe. Nadzorowanie odległe (ang. *distant supervision*) jest to wykorzystanie częściowo strukturyzowanej kolekcji danych, gdzie część kolekcji (strukturyzowana) traktowana jest jako opisująca część drugą (niestrukturyzowaną) [Broś 2013]. Przykładowo, przyjmijmy, że posiadamy kolekcję recenzji, w których recenzenci wyrażali opinię o bytach i ich aspektach na dwa sposoby: w postaci wypowiedzi tekstowej oraz za pomocą oceny w skali punktowej. Zgodnie z ideą nadzorowania odległego, możli-

⁸Jako byt można przyjąć tu określony film, natomiast jako aspekt jego fabułę.

⁹Bytem może być tu samochód, zaś aspektem łatwości kierowania nim. W tym przypadku, zachowanie samochodu na zakręcie przypisano do szerszego aspektu, dla którego system przeprowadza analizę.

we jest przypisanie polarności do sformułowań używanych w recenzjach tekstowych na podstawie ocen w skali punktowej, przypisanych do tych recenzji. Przykładowo, jeśli pewne sformułowanie zazwyczaj pojawia się w wypowiedziach tekstowych w recenzjach, w których nadawane były wysokie oceny punktowe, to prawdopodobnie sformułowanie to wyraża pozytywną opinię.

Cel pracy wpisuje się w opisany nurt nadzorowania odległego. Podejściem, które ma pozwolić na osiągnięcie założonego celu, jest wykorzystanie podsumowań wielu opinii (przez wiele opinii rozumiemy więcej niż jedną opinię) wyrażonych o produktach i usługach do ustalenia polarności sformułowań, używanych w recenzjach tych produktów i usług. W istniejących metodach, dla każdego sformułowania występującego w wypowiedzi tekstowej analizowano wyłącznie strukturyzowane dane pochodzące z tej samej recenzji (tj. ocenę tego samego recenzenta, wyrażoną w przetwarzalny sposób). W proponowanej metodzie wykorzystana będzie analiza współzależności pomiędzy sformułowaniami używanymi w recenzjach wyrażonych w postaci wypowiedzi tekstowych, a znanymi podsumowaniami wielu opinii o produktach, usługach i ich aspektach, których te recenzje dotyczą. W związku z tym, przyjęta w pracy teza brzmi następująco:

Wykorzystanie podsumowania opinii recenzentów o pewnym produkcie lub usłudze, skonstruowanego na podstawie wielu częściowo strukturyzowanych recenzji tego produktu (lub usługi), poprawi skuteczność budowania leksykonu sformułowań wyrażających opinie o różnych jego aspektach, a tym samym pozwoli na opracowanie metody oceny użyteczności produktów i usług przez konsumentów.

Udowodnienie postawionej tezy ma się odbyć poprzez osiągnięcie następujących celów szczegółowych:

1. Analiza możliwości wykorzystania podsumowań wielu opinii o produktach i usługach dla potrzeb ustalania polarności sformułowań wykorzystywanych przez recenzentów do wyrażania opinii.
2. Opracowanie metody identyfikacji polarności sformułowań służących do wyrażania opinii o aspektach produktów i usług na podstawie informacji pozyskanych z podsumowań wielu recenzji konsumenckich, oceniających te produkty i usługi.
3. Opracowanie metody analizy wydźwigniętu wykorzystującej informacje zawarte w podsumowaniach wielu opinii o produktach i usługach.

Uzyskiwane rezultaty badań z informatyki ekonomicznej, zgodnie z pracą [Gregor i Hevner 2013] można zaliczyć do jednej z czterech kategorii, zależnie od dojrzałości istniejących rozwiązań oraz obszaru ich zastosowania, jak zaprezentowano na rysunku 1.1. Te kategorie to:

1. dostarczenie znanego rozwiązania znanego problemu (badania zaliczane do tej kategorii nie wnoszą istotnego wkładu do nauki),
2. wypracowanie nowego rozwiązania znanego problemu,
3. zastosowanie znanego rozwiązania dla rozwiązania nowego problemu,
4. wypracowanie nowego rozwiązania dla nowego problemu.

Badania prezentowane w pracy sytuują się na przecięciu drugiej i czwartej z wymienionych powyżej kategorii. Dotyczą one znanego zagadnienia identyfikacji polarności, jednak jednocześnie zidentyfikowano pewne braki w istniejących podejściach do rozwiązania tego zagadnienia, a tym samym sformułowano nowy problem, polegający na częstym braku dostatecznych informacji do przeprowadzenia identyfikacji polarności na drodze nadzorowania odległego. Praca ma pozwolić na stworzenie rozwiązań, które, na drodze rozwiązania wspomnianego problemu, będą w porównaniu z istniejącymi obecnie podejściami pozwalały na uzyskanie lepszych rezultatów identyfikacji polarności z punktu widzenia trafności (wyrażanej przez miary precyzji i pełności) przypisywania polarności do sformułowań używanych przez recenzentów do wyrażania opinii.

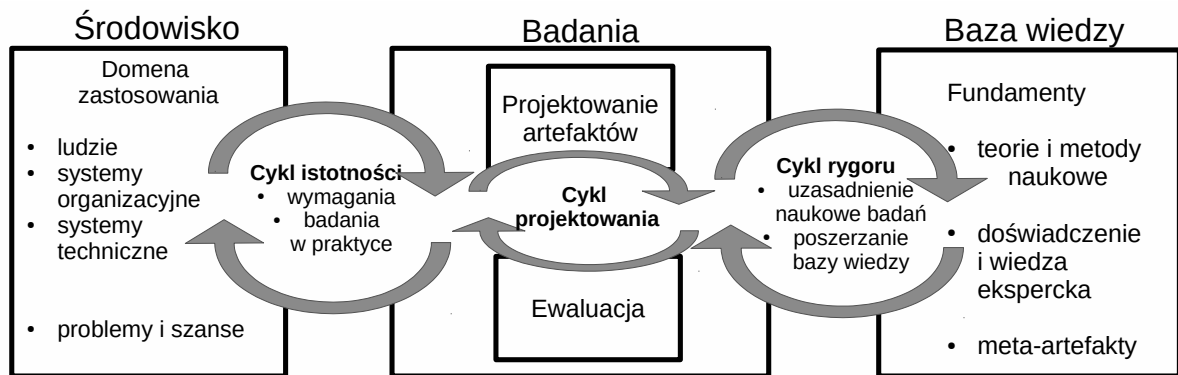
1.3 Metody badawcze i źródła wykorzystywanych informacji

Badania z informatyki ekonomicznej mogą być prowadzone zgodnie z jednym z dwóch paradygmatów:

- nauki behawioralne (ang. *behavioral science*), mające na celu wypracowywanie i weryfikację teorii dotyczących zachowań ludzi i organizacji,
- projektowanie (ang. *design science*), tj. nauki mające na celu rozwiązywanie problemów organizacyjnych przez projektowanie nowych artefaktów, a tym samym poszerzanie zdolności działania ludzi i organizacji [Hevner i in. 2004].

Prezentowana praca zrealizowana została zgodnie z paradygmatem projektowania. Praca ma dostarczyć artefakty, rozwiązujące problem podmiotów będących uczestnikami wymiany rynkowej

polegający na przesycie informacyjnym związanym z nadmiarem opinii konsumenckich publikowanych w Internecie.



Rysunek 1.2: Cykle w procesie badań nad systemami informatycznymi z zakresu projektowania (ang. *design science*). Źródło: [Hevner 2007]

Zakres pracy jest zgodny z modelem zaprezentowanym na rysunku 1.2, zaczerpniętym z pracy [Hevner 2007]. Elementami tego modelu są:

- środowisko (ang. *environment*), czyli pewna domena zastosowania, a wraz z nią należący do niej ludzie, systemy organizacyjne i systemy techniczne oraz występujące w niej problemy i możliwości; analizę tego elementu modelu opisano w rozdziale 2,
- baza wiedzy (ang. *knowledge base*), czyli teorie i metody naukowe, istniejące artefakty oraz udokumentowane wnioski z wcześniejszych działań, które mogą być zastosowane w rozwiązaniu postawionego problemu; tego elementu modelu dotyczy rozdział 3,
- badanie właściwe (ang. *design science research*), na które składają się sprzężone ze sobą dwa elementy: tworzenie artefaktów oraz ich ewaluacja; ten element został opisany w rozdziałach 4, 5, 6 i 7.

Pomiędzy poszczególnymi elementami modelu następują dwukierunkowe przepływy. Przeprowadzone badanie czerpie z analizy otoczenia oraz bazy wiedzy. Jednocześnie, udostępnia ono artefakty możliwe do zastosowania w otoczeniu oraz poszerza istniejącą bazę wiedzy.

Dla tak rozumianego obszaru badawczego, podjęto decyzję o przeprowadzeniu badań zgodnie ze wskazówkami zawartymi w pracy [Österle i in. 2011]. Wyboru dokonano ze względu na zwięzłość i spójność wskazówek zawartych we wspomnianej pracy oraz szeroką ich akceptację w

środowisku naukowym. Proces badań składa się tutaj z czterech głównych faz, odbywających się w wielu iteracjach: analizy, projektowania rozwiązania, ewaluacji i rozpowszechniania wyników. W kolejnych podsekcjach omówiono, jak prowadzone były kolejne fazy badań.

Faza analizy

Faza analizy ma w pracy [Österle i in. 2011] trzy główne cele:

- identyfikację i opisanie problemu biznesowego, którego dotyczy praca oraz wyspecyfikowanie pytań i celów badawczych oraz luk w istniejących rozwiązaniach,
- przeprowadzenie analizy aktualnego stanu wiedzy w dziedzinie,
- stworzenie planu badań, które mają umożliwić wypracowanie założonych artefaktów.

Wyniki tej fazy zostały opisane w rozdziałach drugim, trzecim i czwartym. Znaczną rolę w tej fazie odgrywała przeprowadzona analiza literatury oraz wybranych portali, umożliwiających publikowanie recenzji konsumenckich. Zadania te zostały przeprowadzone z wykorzystaniem dwóch źródeł. Po pierwsze, przeanalizowano niemal 150 publikacji naukowych z zakresu recenzji konsumenckich i ich automatycznego przetwarzania. Na bibliografię pracy składają się publikacje wybrane w następujący sposób:

- wybrano dwie prace [Liu 2011; Pang i Lee 2008], będące najczęściej cytowanymi artykułami zawierającymi przegląd literatury z zakresu analizy wydźwięku (zgodnie z wyszukiwarką Google Scholar, odpowiedź na zapytanie „sentiment analysis survey”¹⁰) i w trakcie ich lektury analizowano abstrakty artykułów cytowanych przez wspomniane dwie prace; spośród nich, na podstawie ich zgodności z celem pracy wybierano dalsze artykuły do analizowania w całości i dodawano je do bibliografii,
- w celu analizy artykułów opublikowanych po obu wspomnianych w poprzednim punkcie pracach, wykorzystano wyszukiwarkę dostępną na stronie biblioteki ACM Digital Library, do której zadano zapytanie „sentiment analysis”, a wyniki ograniczono do artykułów opublikowanych po 2010 roku i posortowano je po relewancji; na drodze analizy abstraktów prac identyfikowano artykuły dotyczące identyfikacji polarności i prezentujące podejścia nowatorskie w porównaniu do uprzednio przeanalizowanych artykułów,

¹⁰jedną z prac zastąpiono inną, dłuższą pracą tego samego autora na ten sam temat, opublikowaną rok wcześniej

- bibliografię uzupełniają dodatkowe prace z Biblioteki Uniwersytetu Ekonomicznego w Poznaniu i znalezione w Internecie, które dotyczyły zagadnień istotnych dla pracy, ale nieporuszanych w przeanalizowanych pracach (np. dotyczących badania satysfakcji klientów czy Web 2.0).

W fazie analizy przeprowadzono również samodzielny przegląd dwudziestu portali (zarówno polskojęzycznych, jak i angielskojęzycznych) umożliwiających publikowanie recenzji konsumenckich w Internecie. Wynikiem analizy była propozycja systematyki takich portali oraz różnych sposobów, na jakie mogą być wyrażane opinie o produktach.

Przeprowadzona w ten sposób analiza umożliwiła stworzenie metody, której główną cechą jest nowatorskie wykorzystanie informacji dostępnych w częściowo strukturyzowanych recenzjach konsumenckich. Metoda szczegółowo została opisana w rozdziale czwartym pracy. Faza ta została dodatkowo poszerzona o eksperymenty, mające na celu analizę wybranego korpusu recenzji konsumenckich w celu oceny możliwości przeprowadzenia wnioskowań założonych w pracy. Wyniki wspomnianych eksperymentów zostały opisane w rozdziale piątym.

Faza projektowania rozwiązania

W tej fazie, zgodnie z wytycznymi z pracy [Österle i in. 2011], ma nastąpić wypracowanie założonych artefaktów za pomocą metod badawczych ogólnie zaakceptowanych dla dziedziny nauki, której badania dotyczą. Dodatkowo, wymagane jest, aby wypracowywane rozwiązanie zostało porównane z istniejącymi rozwiązaniami w celu wykazania oczekiwanych różnic. Istotne jest również, aby dla wypracowanych artefaktów przedstawić odpowiednie uzasadnienie ich istotności dla domeny zastosowania.

Wynikami pracy są cztery rodzaje artefaktów, zgodnie z przyjętymi wskazówkami dla badań projektowych w dziedzinie systemów informatycznych [Gregor i Hevner 2013]:

- terminologia wykorzystywana do opisanie problemów i ich rozwiązań, w tym przypadku związana z recenzjami konsumenckimi publikowanymi w Internecie oraz z analizą wydźwięku, przedstawiona w rozdziałach 2, 3 i 4,
- modele, czyli reprezentacje problemów i ich możliwych rozwiązań, opisujące obiekty wykorzystywane w identyfikacji polarności i analizie wydźwięku; modele te zostały opisane w rozdziałach 4, 5, 6 i 7,

- metody, to jest algorytmy i praktyki postępowania w celu rozwiązywania zdefiniowanych problemów; w pracy zaprezentowano m.in. metody ekstrakcji z tekstu sformułowań służących do wyrażania opinii oraz identyfikacji ich polarności za pomocą podsumowań wielu opinii o produktach i usługach. W wypracowanych metodach wykorzystano m.in. szeroko stosowaną metodę uczenia maszynowego; patrz rozdziały 5, 6 i 7,
- instancje, czyli realizacje modeli i metod działających w świecie rzeczywistym; w trakcie prac przygotowano prototypy będące działającymi implementacjami proponowanych metod (ponownie rozdziały 5, 6 i 7).

Faza ewaluacji

W tej fazie, wypracowane artefakty mają podlegać ewaluacji w celu wykazania osiągnięcia założonych celów. Może się to odbywać m.in. na drodze eksperymentów, symulacji czy ocen dokonanych przez ekspertów [Österle i in. 2011]. W prezentowanej pracy, główną ewaluację wypracowanych metod przeprowadzono na drodze eksperymentów, w których określano prawidłowość ich działania poprzez porównanie wyników działania prototypów z tzw. złotymi standardami, czyli wzorcowymi rezultatami przygotowanymi przez ekspertów. W celu udowodnienia tezy, w dwóch eksperymentach porównano skuteczność proponowanej metody z dwoma istniejącymi podejściami do identyfikacji polarności i analizy wydźwięku.

1.4 Struktura rozprawy

Rozprawa składa się z ośmiu rozdziałów, w skład których wchodzi Wprowadzenie oraz Podsumowanie. Wśród pozostałych sześciu rozdziałów można wyodrębnić trzy główne części. Pierwsza z nich zawiera analizę literatury, natomiast pozostałe dwie części zawierają omówienie wypracowanych w pracy modeli, metod i instancji.

Na pierwszą ze wspomnianych części składają się rozdziały drugi i trzeci. Prezentuje ona wyniki przeprowadzonej analizy literatury i analizy domeny. W drugim rozdziale, zatytułowanym „Recenzje konsumenckie w Internecie i ich rola w podejmowaniu decyzji przez uczestników wymiany rynkowej”, zamieszczona jest analiza zjawiska recenzji konsumenckich publikowanych w Internecie wraz z opisem wpływu takich recenzji na zachowania się podmiotów na rynku. Natomiast rozdział trzeci nosi tytuł „Analiza wydźwięku i metody konstruowania leksykonów sformułowań wyrażających opinie”. Zawiera on wyniki analizy literatury z zakresu automatycz-

nego przetwarzania recenzji konsumenckich, ze szczególnym uwzględnieniem metod mających na celu naukę leksykonów i identyfikację polarności sformułowań wyrażających opinie.

Druga część, składająca się z rozdziałów czwartego i piątego, ma na celu omówienie modeli i metod proponowanych w pracy i przedstawienie planowanego wkładu do dziedziny, a także wstępną analizę możliwości przeprowadzenia proponowanych wnioskowań. W czwartym rozdziale, zatytułowanym „Wykorzystanie podsumowań opinii o produktach i usługach do identyfikacji polarności – zarys podejścia”, wyjaśniono, jakie braki istniejących metod mają na celu naprawić proponowane artefakty oraz opisano główne założenia proponowanej metody, wraz z trzema jej wariantami dopasowanymi do różnych scenariuszy wykorzystania. Rozdział „Analiza spójności różnych sposobów wyrażania opinii o produktach i usługach” prezentuje wyniki eksperymentów, mających na celu wstępną analizę możliwości przeprowadzenia wnioskowań opisanych w rozdziale czwartym. Badania, które zostały w nim przedstawione, miały na celu osiągnięcie pierwszego celu szczegółowego pracy.

Trzecia część pracy szczegółowo opisuje opracowaną metodę, jej implementację oraz jej ewaluację. Obejmuje ona rozdział szósty i siódmy. W rozdziale szóstym, zatytułowanym „Ekstrakcja sformułowań służących do wyrażania opinii i wykorzystanie częściowo strukturyzowanych recenzji konsumenckich w procesie identyfikacji ich polarności” szczegółowo zaprezentowano wypracowaną metodę. Omówione są również jej szczegóły implementacyjne. Rozdział siódmy („Ewaluacja wypracowanej metody”) opisuje wyniki eksperymentów, mających za zadanie wykazanie osiągnięcia celu pracy i weryfikujące jej tezę. Porównane zostały w nim wyniki identyfikacji polarności za pomocą zaprezentowanej metody z wynikami uzyskanymi za pomocą wybranych istniejących metod identyfikacji polarności. Dodatkowo, z wykorzystaniem uzyskanych leksykonów przeprowadzono analizę wydźwiku w celu wykazania większej poprawności tego procesu (w rozumieniu miar precyzji i pełności) w przypadku wykorzystania leksykonu wygenerowanego za pomocą opracowanej metody. W tej części zrealizowane zostały pozostałe dwa (drugi i trzeci) cele szczegółowe pracy.

Praca kończy się podsumowaniem, w którym dokonano przeglądu wypracowanych rezultatów i omówiono, w jaki sposób założone cele pracy zostały osiągnięte.

Rozdział 2

Recenzje konsumenckie w Internecie i ich rola w podejmowaniu decyzji przez uczestników wymiany rynkowej

Celem rozdziału jest przedstawienie wyników analizy literatury z zakresu roli recenzji konsumenckich w podejmowaniu decyzji przez podmioty gospodarcze: zarówno nabywców, jak i sprzedawców. Dodatkowo, w rozdziale omówiono wyniki przeprowadzonej przez autora pracy analizy portali umożliwiających publikowanie takich recenzji.

Rozdział ma następującą strukturę. Najpierw omówiona jest geneza zjawiska recenzji konsumenckich. Następnie, analizie poddane zostało zagadnienie zaangażowania konsumentów w publikowanie recenzji. Omówiono czynniki, które motywują ich do dzielenia się z innymi użytkownikami Internetu swoimi doświadczeniami na temat produktów i usług. W kolejnych sekcjach (2.3 i 2.4) przedstawiono systematykę portali, na których możliwe jest publikowanie recenzji oraz omówiono, na jakie sposoby recenzje mogą być wyrażane. W sekcji 2.5 omówiono natomiast, w jaki sposób recenzje konsumenckie mogą wpływać na zachowania uczestników wymiany rynkowej. Rozdział kończy się krótkim podsumowaniem zawierającym najważniejsze wnioski.

Obszerne fragmenty niniejszego rozdziału zostały opublikowane w pracach [Bernatowicz i Małyszko 2014] i [Małyszko 2013].

2.1 Geneza zjawiska

Internet jest popularnym medium do publikowania opinii o produktach i usługach oraz do zapoznawania się z opiniami opublikowanymi przez innych użytkowników. W czasach przed jego rozpowszechnieniem i we wczesnych fazach jego rozwoju, nabywcy również wymieniali się informacjami o swoich opiniach, jednakże odbywało się to z wykorzystaniem innych sposobów wymiany informacji. W pracy [McAfee 2006] przedstawiono dwa tradycyjne sposoby komunikacji, które można odnieść również do wymiany opinii. Te sposoby to:

- kanały – sposób komunikacji, w którym każdy może być nadawcą treści, jednak krąg odbiorców takich wiadomości jest bardzo ograniczony; przykładem takiego rodzaju komunikacji jest rozmowa telefoniczna, poczta elektroniczna oraz komunikatory internetowe,
- platformy – w tym sposobie komunikacji, wiadomość trafia do bardzo szerokiego grona odbiorców, jednak nadawana może być tylko przez ściśle ograniczone grono nadawców; jako przykład można podać czasopisma, strony intranetowe w korporacjach itp.

W żadnym z tych przypadków nie było możliwe przekazywanie informacji przez dowolną osobę szerokiemu gronu odbiorców [McAfee 2006]. Naturalnymi sposobami wymiany opinii były tu bądź bezpośrednie kontakty wśród ograniczonych liczebnie grup znajomych, bądź też recenzje przygotowywane przez ekspertów w danej dziedzinie, publikowane w specjalistycznych czasopismach lub, później, na stronach internetowych poświęconych określone mu tematowi.

Sytuacja ta uległa zmianie wraz z rozpowszechnieniem się zjawiska Web 2.0, zgodnie z którym rola dostawcy i użytkownika informacji w Internecie stają się coraz trudniejsze do rozgraniczenia [Abramowicz 2008; Lu i in. 2011]. Dzieje się tak z powodu wzrostu zaangażowania użytkowników Sieci w publikowanie informacji z wykorzystaniem takich rozwiązań jak wiki, blogi oraz inne rodzaje systemów zbudowanych zgodnie z zasadami tzw. architektury partycypacji [O'Reilly, T. 2004, 2005, 2006]. Dzięki temu, powstała nowa kategoria recenzji - recenzje konsumenckie (patrz definicja na stronie 1). Za ich pomocą, każdy użytkownik Internetu może opublikować swoją opinię o danym produkcie i usłudze, która to opinia następnie jest dostępna szerokiemu gronu odbiorców.

Opinię nabywcy, opublikowaną w postaci recenzji, można traktować jako wyraz satysfakcji klienta z recenzowanego produktu lub usługi. Zgodnie z normą ISO 9000:200, satysfakcja klienta jest to „percepcja klienta dotycząca stopnia, w jakim jego wymagania są spełnione” [Nieżurawski, Pawłowska i Witkowska 2010]. Satysfakcja klienta jest zjawiskiem bardzo subiektywnym i można

ją traktować jako „wypadkową pomiędzy oczekiwaną jakością produktu, postrzeganą jakością produktu oraz znaczeniem różnych jego właściwości [dla klienta]” ([Baran 2000] za [Niezurawski, Pawłowska i Witkowska 2010]).

2.2 Fenomen zaangażowania konsumentów w publikowanie recenzji

Autorzy recenzji konsumenckich dzielą się z innymi użytkownikami Sieci informacjami i doświadczeniami, związanymi z różnymi produktami i usługami. Motywacja, na podstawie której nabywcy są oni skłonni dzielić się swoimi opiniami, jest interesująca m.in. z socjologicznego punktu widzenia.

Zagadnienie motywacji konsumentów do publikowania swoich opinii w Internecie zostało poddane analizie m.in. w artykule [Bernatowicz i Małyшко 2014]. Zwrócono tam uwagę na fakt, że wspieranie znajomych w podejmowaniu decyzji, tworzenie sieci poradnictwa i wzajemnego polecania jest naturalnym zjawiskiem w gronie znajomych. Autorzy artykułu zwrócili jednak uwagę na fakt, że mechanizmy więziotwórcze, dotyczące zbiorowości tradycyjnych, zostały w pewien sposób przeniesione do Internetu. Co ciekawe, w społecznościach internetowych nie jest konieczne występowanie faktycznej znajomości pomiędzy użytkownikami, aby skłonić ich do wymiany opinii. W socjologicznych analizach relacji między użytkownikami Internetu pojawia się między innymi wątek dotyczący powstawania nowych rodzajów zbiorowości. [Pigla 2012] nazwał takie społeczności *webplemionami*. Społeczności takie powstają w Internecie i łączą ludzi o podobnych zainteresowaniach. Wśród członków takich *webplemion* tworzą się więzi afektualne, co może sygnalizować odczuwaną przez ludzi potrzebę powrotu do tradycyjnego, wspólnotowego charakteru zbiorowości. Ta właśnie potrzeba może tłumaczyć fakt, że użytkownicy Internetu mają tendencję do traktowania innych użytkowników, nawet tych, z którymi nie łączą ich żadne bezpośrednie więzi, jako członków tej samej co oni zbiorowości. W takim kontekście, publikowanie recenzji konsumenckich jest naturalnym przeniesieniem zjawiska występującego w zbiorowościach tradycyjnych na grunt społeczności internetowych [Bernatowicz i Małyшко 2014].

Inne powody zaangażowania konsumentów w publikowanie recenzji podano w pracy [Hennig-Thurau i in. 2004]. Na podstawie przeprowadzonych badań do powodów tych zalicza się:

- chęć pomocy lub zaszkodzenia danemu podmiotowi gospodarczemu (sprzedawcy, producentowi, usługodawcy) poprzez pozytywną lub negatywną recenzję, w zależności od zadowolenia

z jego produktów lub usług nabytych przez recenzenta,

- pobudki altruistyczne – chęć pomocy innym konsumentom (polecenie dobrego produktu lub porada, aby unikać złego), którzy potencjalnie mogą być nabywcami danego produktu lub usługi,
- przyjemność czerpana z samego faktu dzielenia się swoimi przeżyciami,
- publikowanie recenzji bywa nagradzane różnymi zachętami, np. udziałem w konkursie z cennymi nagrodami dla osób zamieszczających recenzje [Barton 2006],
- chęć skłonienia danego podmiotu gospodarczego do podjęcia konkretnych działań związanych z produktem lub usługą, przykładowo poprzez opublikowanie na witrynie podmiotu negatywnej opinii tego produktu lub usługi, wskazującej na wykryte wady.

2.3 Portale udostępniające możliwość publikowania recenzji

Opinie nabywców na temat produktów i usług mogą być publikowane na różnego rodzaju portalach internetowych. W tej sekcji opisano wypracowaną systematykę typów witryn internetowych, na których najczęściej takie recenzje są publikowane, oraz wskazując na ich cechy charakterystyczne. Prezentowaną systematykę wypracowano na podstawie analizy portali udostępniających możliwość publikowania recenzji konsumenckich. Spośród wielu przeanalizowanych portali wybrano 20 będących reprezentatywnymi przykładami portali o podobnych cechach. Dla wybranych portali przeprowadzono pogłębioną analizę, której wynikiem jest przedstawiona systematyka. Do skonstruowania listy tych portali wykorzystano wyszukiwarkę internetową Google oraz katalog podobnych portali pl.similarities.com. Przy ustalaniu listy kierowano się następującymi dwiema zasadami:

- szukano polskich odpowiedników dla portali zagranicznych omawianych w literaturze,
- szukano portali o różnorodnych charakterystykach tak, aby ukazać szeroki przegląd podejść do publikowania recenzji konsumenckich.

W trakcie analizy, zidentyfikowano dziewięć typów portali umożliwiających zamieszczanie i publikowanie recenzji. Poniżej znajduje się omówienie wypracowanej systematyki.

- **Portale tematyczne.** Portale takie są wyspecjalizowane w kierunku opisywania konkretnego typu produktów lub usług, przykładowo książek (lubimyczytac.pl), filmów (filmweb.pl),

restauracji (gastronauci.pl), hoteli (tripadvisor.com) itp. Zakres tematyczny takich portali może mieć różną szerokość - przykładowo, na stronie gastronauci.pl znajdują się recenzje tylko lokali gastronomicznych, natomiast na portalu tripadvisor.com znajdują się recenzje takich właśnie lokali oraz dodatkowo hoteli i atrakcji turystycznych. Dla użytkowników portale tego typu zazwyczaj mają stanowić źródło szczegółowej wiedzy o opisywanej dziedzinie i, oprócz recenzji, mogą zawierać takie treści, jak informacje o najnowszych wydarzeniach i wywiady z osobami ważnymi dla danej dziedziny (np. autorami książek na portalu lubimyczytac.pl).

- **Ogólnotematyczne portale z recenzjami.** Pod tym mianem rozumiemy portale poświęcone wyłącznie recenzowaniu produktów, przy czym zakres recenzowanych produktów i usług jest tu bardzo szeroki (nie skupiają się one na żadnym konkretnym ich rodzaju). Na rysunku 2.1 zaprezentowano fragment drzewa kategorii produktów recenzowanych na portalu cokupic.pl, obejmujący bardzo szeroki zakres produktów. W Polsce dużymi portalami tego typu są między innymi wspomniany cokupic.pl, opineo.pl (w którym można zamieszczać również opinie o sklepach) oraz znam.to.

Mapa kategorii

<u>Biuro i firma</u> (377958)	<u>Gry</u> (59566)
<u>Biżuteria i zegarki</u> (256017)	<u>Hobby</u> (462821)
<u>Dla dziecka</u> (313619)	<u>Komputery</u> (1305374)
<u>Dom i wnętrze</u> (2582790)	<u>Księgarnia</u> (3795951)
<u>Erotyka</u> (108704)	<u>Motoryzacja</u> (3815286)
<u>Filmy</u> (67696)	<u>Muzyka</u> (434081)
<u>Fotografia</u> (168408)	<u>Odzież, obuwie, dodatki</u> (1161380)

Rysunek 2.1: Fragment pierwszego poziomu drzewa kategorii produktów na portalu cokupic.pl. Obok nazw kategorii wypisane są liczby produktów znajdujących się w tych kategoriach (stan na dzień 4 września 2012). Źródło: cokupic.pl

- **Porównywarki cenowe.** Portale tego typu mają za zadanie agregację informacji o cenach produktów i usług w sklepach i zakładach usługowych. Na niektórych portalach tego typu

możliwe jest zamieszczanie recenzji zarówno produktów i usług, dla których ceny są porównywane, jak również sklepów i zakładów, z których porównywarka czerpie informacje o cenach. Przykładami takich portali są ceneo.pl i skapiec.pl.

- **Sklepy internetowe.** Coraz częściej, sklepy prowadzące swą działalność przez Internet dają użytkownikom możliwość nadania oceny poszczególnym produktom i usługom przez nie oferowanym. Jednym z pionierów tego typu rozwiązań był sklep amazon.com. Portalami tego typu w Polsce są m.in. merlin.pl oraz komputronik.pl.
- **Portale aukcyjne.** Na portalach aukcyjnych zazwyczaj funkcjonują systemy oceniania sprzedających i kupujących mające na celu nieść pomoc użytkownikom w określeniu wiarygodności partnera transakcji. Po wykonanej transakcji, strony wyrażają opinie o sobie i o jakości dokonanej transakcji. Opinie dotyczyć mogą zgodności przedmiotu z opisem, kontaktu ze sprzedającym, czasu realizacji zamówienia itp. Taki system ocen funkcjonuje m.in. na portalu allegro.pl.
- **Strony producentów i usługodawców.** Możliwość oceny udostępniana bywa również na stronach producentów i usługodawców, prezentujących informacje o swoich produktach i usługach (np. na portalu www.panasonic.co.uk), gdzie klienci, posiadający produkty firmy, proszeni są o wyrażenie o nich opinii lub też opinii o całej firmie. Czasem ta funkcja jest integralną częścią firmowego sklepu internetowego. Wówczas ta kategoria witryn jest podobna od opisywanej wcześniej kategorii sklepów internetowych. Należy zwrócić uwagę na fakt, że często firmy udostępniają na stronie możliwość oceny produktów czy usług wyłącznie w postaci ankiety, której wyniki nie są publikowane. Na innych stronach publikowane są jedynie wyselekcjonowane opinie nadesłane przez użytkowników.
- **Fora internetowe.** Funkcjonuje wiele forów, skupiających osoby zainteresowane konkretnym tematem (np. książkami, motoryzacją itp.). Fora takie często funkcjonują przy większych portalach poświęconych danemu zagadnieniu. Wśród dyskusji na forach często pojawiają się również wypowiedzi recenzujące produkty i usługi. Na forach mogą istnieć osobne działy służące zamieszczaniu recenzji (np. dział Recenzje Książek na forum ebooki.linuxpl.com). Istnieją również fora w całości poświęcone publikowaniu recenzji konsumentów (np. www.forum.filmowo.net). Zazwyczaj wątek na takim forum rozpoczyna się od recenzji, pod którą nawiązuje się dyskusja dotycząca recenzji i jej przedmiotu. Warto wspomnieć, że funkcja zbliżona do forum może być również dostępna na wielu innych typach

portali (np. na portalach tematycznych) jako dodatkowa możliwość wypowiedzi użytkowników (np. dyskusje na stronach filmów na portalu filmweb.pl).

- **Portale społecznościowe.** Możliwość komunikacji ze znajomymi i publikowania wypowiedzi (z wykorzystaniem usługi tzw. mikroblogowania), będąca istotną cechą portali społecznościowych, również może być wykorzystywana do wyrażania opinii o produktach i usługach [Maynard, Bontcheva i Rout 2012]. Specyficzną cechą wielu portali społecznościowych jest brak anonimowości. Treści, w tym opinie, publikowane są tu przez znane z imienia i nazwiska osoby, które dodatkowo często są znajomymi osób, które się z tymi treściami zapoznają. Może to wywierać większy wpływ tak opublikowanych opinii na ich odbiorców niż w przypadku recenzji nieznanych, anonimowych osób. Jednocześnie, ilość różnorodnych treści pojawiających się na portalach społecznościowych znacząco utrudnia wyszukiwanie publikowanych na nich recenzji. W takim przypadku, zazwyczaj użytkownicy portali społecznościowych zapoznają się z recenzjami niejako „przy okazji”, przykładowo wówczas, gdy natkną się na niedawno opublikowaną wypowiedź znajomego, wyrażającą opinię o określonym produkcie czy usłudze. Istnieją również portale społecznościowe udostępniające specjalne funkcje ułatwiające publikowanie recenzji konsumenckich i zapoznawanie się z nimi. Przykładowo, na portalu facebook.com istnieje możliwość publikowania recenzji na profilu restauracji w postaci oceny punktowej i wypowiedzi tekstowej.
- **Blogi.** Nabywcy mogą dzielić się swoimi opiniami także na blogach. Istnieją wyspecjalizowane blogi dotyczące recenzji określonych rodzajów produktów i usług. Kategoria ta znacznie różni się od pozostałych omawianych kategorii, ponieważ na jednym blogu liczba osób publikujących recenzje jest z reguły bardzo niewielka (jest to jedna lub kilka osób).

Oprócz przynależności do różnych grup w ramach zaproponowanej systematyki, portale umożliwiające publikowanie recenzji konsumenckich różnią się od siebie dodatkowymi cechami. Warto zwrócić uwagę na zakres recenzowanych przedmiotów w poszczególnych portalach. Oprócz tego, że może być on szerszy (np. sklepy i towary z najróżniejszych kategorii, jak na opineo.pl) albo węższy (np. wyłącznie książki na lubimyczytac.pl), charakterystyczny jest tutaj również fakt zamknięcia lub otwarcia listy recenzowanych bytów na modyfikację ze strony użytkowników. Przykładowo, na forach internetowych (forum.filmowo.pl oraz ebooki.linux.pl) to użytkownicy zakładają wątki poświęcone konkretnym recenzowanym filmom czy książkom. Podobnie, na portalu filmweb.pl użytkownicy mogą dodać film do bazy a następnie napisać jego recenzję. Z

drugiej strony, na wielu portalach (takich jak porównywarki cenowe, sklepy internetowe czy strony producentów) takiej możliwości nie ma – byty podlegające recenzowaniu są ustalane wyłącznie przez redakcję portalu.

Omawiane portale różnią się między sobą również tym, że na niektórych z nich, oprócz opinii konsumenckich, zamieszczane są również recenzje profesjonalne, przygotowane przez krytyków lub redakcję danego portalu. Ten drugi przypadek ma często miejsce w portalach tematycznych (np. `filmweb.pl` i `lubimyczytac.pl`) oraz ogólnych portalach z recenzjami (`cokupic.pl`). Na takich portalach, oba typy recenzji są odpowiednio wyróżnione, na przykład umieszczane są w osobnych działach portalu.

Ciekawą obserwacją stanowi fakt, że na niektórych portalach może mieć miejsce weryfikacja zamieszczanych tam recenzji przez redakcję danego portalu. Przykładowo, według informacji dla restauratorów zawartych na stronie `gastronauci.pl`, wszystkie recenzje zamieszczane na tej stronie są analizowane pod kątem ich konstruktywności. Również, w przypadku fałszywości informacji zawartych o lokalu, która jest możliwa do obiektywnego zweryfikowania, właściciel lokalu może poprosić o usunięcie takich opinii. Jednocześnie, weryfikacji takiej mogą dokonać również inni użytkownicy, gdyż nieraz również same recenzje mogą podlegać ocenie innych użytkowników, jak to ma miejsce na przykład na portalu `filmweb.pl`, gdzie czytelnicy mogą stwierdzić, czy uważają daną recenzję za „pomocną”. Takie możliwości weryfikacji recenzji pełnią coraz istotniejszą rolę ze względu na zjawisko nieuczciwych zachowań przedsiębiorstw, na przykład mających na celu zamieszczanie fałszywych, niepoehlebnych recenzji produktów lub usług konkurencji lub nieprawdziwych opinii dotyczących oferowanych przez przedsiębiorstwo produktów czy usług. Kwestia problemów z wiarygodnością recenzji zostanie szerzej omówiona w sekcji 2.5.2.

Podsumowując warto wspomnieć, że portale umożliwiające publikowanie opinii często posiadają funkcję systemów rekomendacyjnych. Na podstawie ocen, jaki dany użytkownik nadaje recenzowanym przedmiotom, system taki może przewidzieć, jakie prawdopodobnie oceny nadałby innym obiektom, dzięki czemu może mu polecić te, które najprawdopodobniej najbardziej przypadną mu do gustu [Su i Khoshgoftaar 2009].

2.4 Formaty wyrażania opinii

W poprzedniej sekcji omówiono systematykę portali, na których można znaleźć opinie wyrażane przez nabywców na temat produktów i usług. W tej sekcji omówione zostaną sposoby wyrażania

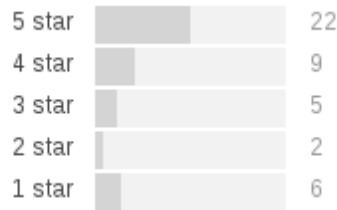
opinii (będziemy je nazywać formatami wyrażania opinii). Należy zwrócić uwagę na fakt, że opisane poniżej formaty często są ze sobą łączone. Opinia może dotyczyć całego produktu lub poszczególnych jego cech czy elementów, które będziemy nazywać aspektami¹.

Jak wspomiano, istotnym zagadnieniem dla niniejszej pracy jest możliwość podsumowywania wielu opinii wyrażanych o dobrach i ich aspektach. W związku z tym, poniżej omówiono również możliwości podsumowywania opinii wyrażanych na poszczególne sposoby.

- **Skala punktowa.** Na wielu portalach (przykładowo `cokupic.pl` oraz `amazon.com`) recenzenci mogą przypisywać do produktów oceny punktowe. Mogą one być przypisywane do całego produktu lub jego aspektów, gdzie lista takich aspektów jest zdefiniowana dla danego produktu (przykładowo, na portalu `gastronauta.pl` takie aspekty dla restauracji to: jedzenie, wystrój, obsługa oraz jakość/ceny). Popularną reprezentacją graficzną przyznawania ocen w ten sposób są tzw. „gwiazdki”, dzięki którym użytkownik może nadać ocenę w skali od 1 do 5 klikając na odpowiednią gwiazdkę [Hu, Zhang i Pavlou 2009]. Ważną zaletą takiego sposobu oceniania jest jego szybkość i intuicyjność, a także łatwość przeprowadzania późniejszych analiz na tak nadanych ocenach. Przykładowo, określenie średniej oceny danego produktu wymaga jedynie policzenia średniej arytmetycznej przyznanych punktów. Podejście takie często traktowane jest jako ogólna ocena danego produktu lub jego aspektu. Metoda ta stosowana jest w większości portali z recenzjami oraz sklepów internetowych, np. w systemie `cokupic.pl` oraz w sklepie `amazon.com`. Niestety, istniejące wyniki badań wskazują, że średnia punktów może być mylnym wskaźnikiem sposobu postrzegania dóbr przez recenzentów, choćby z uwagi na fakt bimodalności rozkładu ocen (często dla jednego produktu recenzenci przyznali wiele bardzo pozytywnych, jak i jednoznacznie negatywnych ocen) [Hu, Pavlou i Zhang 2006]. Innym sposobem podsumowania opinii wyrażonych w ten sposób jest określenie rozkładu punktów względem liczby recenzentów, którzy nadali produktowi (lub jego aspektowi) daną ocenę (patrz rysunek 2.2). Kolejną potencjalnie przydatną informacją dla użytkownika, możliwą do wygenerowania na podstawie średniej liczby punktów przyznanych danemu produktowi, może być miejsce w rankingu produktów, ustalonym na podstawie średniej punktacji, co ma miejsce na portalu `filmweb.pl`.

- **Czy polecasz ten produkt?** Jest to najbardziej ogólny format wyrażenia opinii, w którym

¹Definicja aspektu podana została we Wprowadzeniu do pracy; bardziej szczegółowa analiza różnych definicji aspektu znajduje się w sekcji 3.2.1



Rysunek 2.2: Podsumowanie wielu ocen punktowych w postaci rozkładu liczby recenzentów, którzy przypisali produktowi określoną liczbę punktów. Źródło: amazon.com

recenzent określa jedynie to, czy produkt wart jest polecenia. Metoda ta może być traktowana jako szczególny przypadek skali punktowej z tylko dwiema możliwymi wartościami oceny – pozytywną (w sytuacji rekomendowania produktu) lub negatywną (w przeciwnym wypadku). Ten typ oceny dostępny jest przykładowo na portalu cokupic.pl. Podobnie, oceniając sprzedawcę w portalu Allegro, podstawowym sposobem oceny kontrahenta jest wyrażenie opinii dotyczącej ogółu transakcji (pozytywny, negatywny lub neutralny). Ten typ oceny umożliwia skonstruowanie prostej statystyki, informującej użytkowników jaki odsetek oceniających poleca ten produkt lub usługę (patrz rysunek 2.3).

- Listy zalet i wad.** Często opinie mogą być wyrażane w postaci list zalet i wad. Wartości w każdej z tych list mogą być albo dowolne, jak ma to miejsce na przykład na portalu bozo.pl (gdzie użytkownik ma możliwość wprowadzenia dowolnego ciągu znaków) lub wybierane z predefiniowanej dla danej kategorii produktów listy. Przykładowo, na stronie cokupic.pl w kategorii Gry komputerowe / Gry PC, zalety to m.in. „łatwość nauczenia się”, „intrygująca fabuła” oraz „dobra jakość dźwięku” (w sumie jest tu 11 możliwych do opisanego zalet), zaś jako wady można wybrać takie pozycje jak „nudny”, „brak możliwości grania on-line” i „mało porad” (w sumie zdefiniowanych jest 8 wad). Możliwe jest również rozwiązanie mieszane, gdzie użytkownicy mają do dyspozycji określoną pulę zalet i wad, które mogą przypisać do produktu lub usługi poprzez wybranie ich z gotowej listy, ale mogą też do takiej listy dodawać własne propozycje (jak ma to miejsce przykładowo na stronie www.panasonic.co.uk). Podsumowanie opinii wyrażonych w tym formacie za pomocą predefiniowanych zalet i wad może być łatwo wygenerowane, przykładowo poprzez określenie liczby recenzentów, którzy wybrali daną zaletę lub daną wadę w recenzji produktu. Dzięki temu, czytelnicy mogą szybko poznać najczęściej wymieniane wady i zalety danego produktu (patrz rysunek 2.3).

90% Użytkowników poleca The Sims 2 (Gra PC)

Użycie

- dla nastolatków (1396)
- dzieci (1072)
- dorośli (845)

+ Plusy:

- łatwość nauczenia się (69%)
 - grywalność (69%)
 - grafika (68%)
 - dobra sterowność (40%)
 - intrygująca fabuła (33%)
- [więcej...](#)

- Minusy:

- brak możliwości grania on-line (25%)
 - nudny (21%)
 - brak multiplayera (19%)
 - kiepska fabuła (14%)
 - stara (10%)
- [więcej...](#)

Rysunek 2.3: Fragment podsumowania recenzji gry komputerowej The Sims 2 na portalu cokupic.pl. Znajduje się tu m. in. informacja o tym, jaki procent recenzujących poleca tę grę oraz podsumowanie zaznaczonych przez użytkowników zalet i wad danego produktu. Źródło: <http://cokupic.pl/produkt/The-Sims-2-Gra-PC>. Dostęp: 11 marca 2013

Jednak w przypadku wariantu tego formatu, w którym poszczególne listy zalet i wad mają postać wypowiedzi tekstowej, podsumowywanie recenzji staje się trudniejszym zadaniem i wymaga wykorzystania technik przetwarzania języka naturalnego.

- **Wypowiedź tekstowa.** Przez ten format wyrażania opinii rozumiemy tekst w języku naturalnym, bez ustalonej struktury, w którym autor opinii wyraża swoje zdanie na temat recenzowanego obiektu. Takie opisy mogą mieć różną długość, mogą być napisane różnym stylem, skupiać się na różnych aspektach recenzowanych bytów itd. W ramach pojedynczej recenzji na danym portalu ten sposób wyrażania opinii może być wykorzystany w kilku polach (przykładowo jako krótki tytuł podsumowujący opinię oraz długi tekst szczegółowo wyrażający opinię autora, jak ma to miejsce na portalu amazon.com). Dokładniejsza analiza tego sposobu wyrażania opinii umieszczona została w sekcji 2.4.1 na stronie 30 oraz w

rozdziale 3.

- **Inne.** Recenzje mogą być zamieszczane również w innych formatach. Przykładowo, coraz popularniejsze są recenzje w postaci krótkich filmów, zamieszczanych przykładowo na portalu [youtube.com](https://www.youtube.com). Również sklep [amazon.com](https://www.amazon.com) umożliwia zamieszczanie recenzji w postaci filmów (o maksymalnej długości 10 minut). Ten rodzaj recenzji wykorzystywany jest jednak częściej w przypadku recenzji eksperckich, przykładowo w serwisie [cnet.com](https://www.cnet.com). Analiza tego typu formatów wyrażania opinii pozostaje poza obszarem zainteresowania pracy.

Warto zwrócić uwagę na fakt, że recenzje często zawierają nie tylko opinie, ale też dodatkowe informacje, które nie mają na celu oceny produktu, a jedynie wzbogacenie informacji zawartej w recenzji. Tego typu dodatkowymi informacjami są m. in.:

- informacje o recenzującym mogące pomóc czytelnikowi w ustaleniu tego, z jakiego punktu widzenia produkt lub usługa były oceniane. Przykładowo, na portalu [cokupic.pl](https://www.cokupic.pl) recenzujący mogą opisać samych siebie na drodze zaznaczenia pewnych predefiniowanych pól (np. „technicznie doświadczony”, „zapalony gracz”, „pracowity biznesmen” w kategorii „Laptopy”),
- użytkownicy mogą uzupełniać recenzje o informacje o tym, jaki ich zdaniem może być najbardziej odpowiedni sposób użycia danego produktu. Na portalu [cokupic.pl](https://www.cokupic.pl), recenzenci mają możliwość oznaczenia przedmiotu recenzji poprzez wybór odpowiedniej pozycji z predefiniowanej listy. Przykładowo, w kategorii laptopów takimi sposobami mogą być „w podróży”, „biznes”, „do gier”, „muzyczne” itp. Z kolei na portalu [tripadvisor.com](https://www.tripadvisor.com) recenzje hoteli są podzielone na napisane z punktu widzenia rodzin z dziećmi, par, pojedynczych osób lub osób będących w podróży służbowej,
- obiektywne informacje dotyczące recenzowanego przedmiotu. Istnieją domeny, gdzie pewne informacje o recenzowanych produktach i usługach nie są łatwo dostępne, a ich zamieszczanie w recenzjach jest bardzo przydatne dla potencjalnych czytelników. Przykładowo, w pracach [Dzikowski i in. 2011; Kaczmarek i in. 2012] zaprezentowano aplikację mobilną, służącą recenzowaniu restauracji, ich wyszukiwaniu oraz rekomendacji. Oprócz m.in. opinii wyrażanej o restauracjach za pomocą liczby punktów przypisywanych poszczególnym ich aspektom, możliwe było tu dodawanie takich informacji jak typ kuchni serwowanej w restauracji, oferowane atrakcje (przykładowo, muzyka na żywo bądź transmisje sportowe) czy też możliwość płatności kartą. Oczywiście, opisane powyżej rodzaje informacji mogą być

zawarte w opisie tekstowym, ale portale mogą również udostępniać specjalne mechanizmy służące ułatwieniu dodawania takich informacji w strukturyzowanej formie.

Podsumowanie najważniejszych z punktu widzenia pracy cech poszczególnych formatów za-prezentowano w tabeli 2.1. Poszczególne formaty omówiono tam z punktu widzenia:

- zakresu ocenianych aspektów przedmiotu – w jakim stopniu możliwe jest wyrażanie przez recenzentów opinii o aspektach danego produktu lub usługi na różnych poziomach szczegółowości,
- intensywności opinii – czy możliwe jest nadawanie wyrażanej opinii różnych stopni intensywności,
- stopnia trudności podsumowywania ocen – na ile skomplikowane jest automatyczne przetwarzanie opinii wyrażanych w tym formacie, przykładowo dla potrzeb generowania ich podsumowań.

Format oceniania	Zakres ocenianych aspektów przedmiotu	Intensywność opinii	Stopień trudności podsumowywania ocen
skala punktowa	ściśle ustalony przez redakcję portalu i zazwyczaj wąski; zwykle pojedyncza ocena całego przedmiotu lub kilka ocen dla wybranych aspektów przedmiotu	uzależniony od skali punktowej zdefiniowanej przez redakcję portalu	niski (podsumowywanie łatwe); np. proste wyliczanie średnich lub rozkładów dla ocenianych aspektów
czy polecasz ten produkt?	ocenie podlega wyłącznie cały produkt (brak rozróżnienia na poszczególne aspekty)	binarna wartość (produkt polecany lub odradzany), ew. brak oceny	niski; proste wyliczenie głosów polecających i odradzających
listy zalet i wad	szeroki; listy możliwych wad i zalet mogą mieć wiele pozycji. Na wielu portalach listy te mogą być rozszerzane przez użytkowników, na niektórych nie istnieje predefiniowana lista możliwości, a użytkownik samodzielnie definiuje wady i zalety (krótkie wypowiedzi w języku naturalnym)	binarna wartość: dany aspekt występuje na liście wad lub na liście zalet, ew. brak oceny danego aspektu; intensywność opinii może być wyrażona w wariancie, gdy pozycje listy mają postać wypowiedzi tekstowych	niski lub wysoki, w zależności od tego, czy lista możliwych wad i zalet jest predefiniowana, czy też wprowadzana przez każdego użytkownika osobno; narzucona struktura (krótkie wypowiedzi o znanym wydźwięku) ułatwia przetwarzanie
wypowiedź tekstowa	nieograniczony; recenzent może odnosić się do dowolnych aspektów recenzowanego przedmiotu za pomocą dowolnych sformułowań	dowolna; za pomocą języka naturalnego recenzent może wyrazić swoją opinię o dowolnym stopniu intensywności	bardzo wysoki; brak narzuconej struktury wypowiedzi i wieloznaczność języka naturalnego wymagają wykorzystania zaawansowanych technik przetwarzania języka naturalnego

Tabela 2.1: Porównanie różnych formatów wyrażania opinii. Źródło: opracowanie własne

2.4.1 Wypowiedź tekstowa jako sposób wyrażenia opinii

Szczególnie interesującym formatem wyrażania opinii z punktu widzenia informatyki jest wypowiedź tekstowa. Warto w tym miejscu dokładniej poddać analizie ten sposób wyrażania opinii.

Jako opinię wyrażoną w postaci wypowiedzi tekstowej rozumiemy tekst w języku naturalnym, bez narzuconej struktury, w którym autor opinii wyraża swoje zdanie na temat recenzowanego obiektu. Czasami takie opisy są jedynym sposobem, w jakim opinia zostaje wyrażona (ma to miejsce np. w przypadku recenzji zamieszczanych na forach internetowych). Recenzje konsumenckie w postaci wypowiedzi tekstowej mogą być różnej długości, być pisane różnym stylem, skupiać się na różnych aspektach recenzowanych przedmiotów itp. Zależać to może między innymi od kultury przyjętej na danym portalu czy też domeny produktów podlegających ocenie, na co zwrócono uwagę m.in. w pracy [Gamon i in. 2005]. Autorzy zauważają duże różnice pomiędzy recenzjami filmów i samochodów z punktu widzenia poprawności gramatycznej, liczby błędów ortograficznych itp. Na niektórych portalach można napotkać wiele recenzji tekstowych, które są ledwie kilkudziesięcioma wypowiedziami napisanymi z wykorzystaniem kolokwialnego języka. Na innych portalach takie niewyszukane recenzje mogą się spotkać z krytyką innych użytkowników, bądź też z usunięciem przez moderatora. Przykładowo, zgodnie z regulaminem serwisu [gastronauci.pl](http://www.gastronauci.pl), recenzje, które zostaną uznane za nieprzedstawiające wartości dla portalu, nie są uwzględniane przy obliczaniu średniej oceny lokalu ².

2.5 Recenzje konsumenckie a zachowania podmiotów na rynku

Dalsza część rozdziału zostanie poświęcona analizie, w jaki sposób recenzje konsumenckie wpływają na zachowania podmiotów na rynku. Przeanalizowane zostaną zagadnienia wpływu recenzji zarówno na konsumentów, podejmujących decyzje zakupowe, jak i na sprzedawców, którzy mogą przyjmować różne strategie reagowania na to zjawisko.

2.5.1 Wpływ recenzji na decyzje zakupowe konsumentów

Recenzje konsumenckie mogą stanowić dla konsumentów cenne źródło informacji o produktach i usługach. Z punktu widzenia ekonomii i teorii popytu, wybór określonego koszyka towarów przez konsumenta zależy od jego dochodów (ograniczenia budżetowego), cen towarów i jego gustów. Gusta są tu przyjęte jako dane i nie są głębiej analizowane. Reprezentuje się je za pomocą relacji

²<http://www.gastronauci.pl/pl/artykuly/1612-regulamin>. Dostęp: 17 stycznia 2013

preferencji [Panek 2000]. W tym podstawowym modelu, wpływ recenzji konsumenckich można obrazować w postaci zmian relacji preferencji. Przykładowo, często może zaistnieć sytuacja, gdy dla dwóch inaczej nierozróżnialnych z punktu widzenia konsumenta koszyków, preferuje on ten wyżej oceniany przez innych konsumentów lub ten, w którym wyżej oceniane są najważniejsze dla danego konsumenta aspekty.

W pracy [Rudnicki 2004] przedstawiono bardziej szczegółowy model uwarunkowań zachowań konsumentów. Wyróżniono tu cztery główne grupy takich uwarunkowań:

- uwarunkowania personalno-demograficzne (takie jak wiek, płeć, wykształcenie itd.),
- uwarunkowania psychologiczne (np. nawyki, styl życia, skłonność do ryzyka),
- uwarunkowania ekonomiczne (dochody i ceny, produkty, reklama itp.),
- uwarunkowania społeczno-kulturowe (grupy odniesienia czy liderzy opinii, będące wzorcami do naśladowania lub źródłami opinii).

Kierowanie się recenzjami konsumenckimi można uznać za przejaw uwarunkowań społeczno-kulturowych. Zgodnie z nimi, ludzie z reguły kierują się postawami lub radami innych w przypadku podejmowania decyzji konsumenckich. Za przykład takich uwarunkowań może posłużyć koncepcja tzw. liderów opinii, rozumianych jako osoby, które świadomie lub nieświadomie wpływają na zachowania i decyzje konsumpcyjne otoczenia [Rudnicki 2004]. Potencjalni nabywcy mogą zwracać się do nich z prośbą o opinię przed dokonaniem zakupu. Rola taka jest szczególnie ważna, gdy zakup związany jest z dużym ryzykiem dla konsumenta (np. kupno po raz pierwszy nieznanego produktu o wysokiej cenie) [Chakravarthy i Prasad 2011; Rudnicki 2004]. Postępowanie zgodnie z uzyskaną radą zmniejsza ryzyko podjęcia błędnej decyzji. W przypadku recenzji konsumenckich mamy do czynienia z podobnymi pobudkami, jednak w tym przypadku występuje nie opinia pojedynczego lidera, lecz wyrażona przez wiele osób duża liczba opinii, z którymi użytkownik może się zapoznać przed podjęciem decyzji.

W pracy [Bickart i Schindler 2001] wymieniono trzy fakty natury psychologicznej, które sprawiają, że recenzje konsumenckie mogą szczególnie mocno wpływać na zachowania konsumentów:

- autorzy takich recenzji teoretycznie nie mają żadnego interesu w przekazywaniu fałszywych informacji,
- komunikat otrzymany od osoby, którą łączy pewne podobieństwo z odbiorcą, jest przez

odbiorcę odbierany jako bardziej wiarygodny; w tym przypadku takie podobieństwo polega na tym, że zarówno nadawca, jak i odbiorca są konsumentami,

- recenzje konsumenckie często zawierają opisy sytuacji i doświadczeń związanych z opisywanym przedmiotem; komunikaty zawierające takie elementy często silniej przyciągają uwagę czytelnika i wywołują u niego poczucie empatii z autorem, w ten sposób mocniej przekazując opisywane emocje.

Liczne przeprowadzone w ostatnich latach badania potwierdzają, że przy podejmowaniu decyzji konsumenci przywiązują coraz istotniejszą wagę recenzjom konsumenckim zamieszczanym w Internecie. W pracy [Pang i Lee 2008] wyliczono 30 prac, w których wykazano istotny efekt ekonomiczny tzw. polarności recenzji, czyli binarnych wartości odpowiadających opiniom pozytywnym lub negatywnym, przypisywanym recenzowanym produktom lub usługom. Jednocześnie, wyliczono również siedem prac, które nie pozwoliły na potwierdzenie takiej zależności [Pang i Lee 2008].

Spośród opublikowanych badań warto przywołać następujące wyniki:

- 70% użytkowników internetu ufa opiniom nieznanym im osób zamieszczonym w Internecie, a dla opinii znajomych ta wartość osiąga 90%. Obie te wartości są wyższe od pozostałych form nakłaniania do zakupów (np. reklamy) [The Nielsen Company 2009],
- wśród osób planujących podróże, które wykorzystują Internet w trakcie planowania, od 25% do 33% ankietowanych zdecydowało się na zmianę hotelu pod wpływem opinii innych konsumentów przeczytanych w Sieci [Gretzel i Yoo 2008],
- prawie co czwarty ankietowany użytkownik Internetu stwierdził, że konsultował się z recenzjami konsumenckimi w Sieci również przed dokonaniem zakupu offline, a za produkty ocenione wysoko (na 4-5 gwiazdek) konsumenci są skłonni płacić średnio 20% więcej [comScore 2007].

Należy również zwrócić uwagę na fakt, że w świetle przeprowadzonych badań, stosunek konsumenta do recenzji może być różny w zależności od jego cech (np. demograficznych). Znaczenie tu mają takie czynniki jak doświadczenie z korzystania z Internetu [Zhu i Zhang 2010] oraz wiek [Gretzel i Yoo 2008]. Dodatkowo, recenzje mają różny wpływ na konsumentów w zależności od przedmiotu, którego dotyczą. W pracy [Gretzel i Yoo 2008] stwierdzono, że konsumenci planujący

podróż wakacyjną przywiązują dużo większą wagę do opinii na temat hoteli niż restauracji, atrakcji czy też okresu, w którym najlepiej się wybrać w podróż. Na tej podstawie można przypuszczać, że istnienie badań wskazujących zarówno dużą korelację, jak i jej brak, może wynikać z przeprowadzania ich na różnych danych, reprezentujących różne produkty czy różnych użytkowników Sieci [Zhu i Zhang 2010].

2.5.2 Reakcje sprzedawców na fenomen recenzji konsumenckich dostępnych w Sieci

Podmioty zajmujące się sprzedażą produktów i usług w różnym tempie dostrzegały możliwości, jakie wiążą się z recenzjami konsumenckimi. Wśród pionierów wykorzystania recenzji należy wymienić portale Amazon.com oraz eBay [Barton 2006]. Wielu sprzedawcom uświadomienie sobie potencjału recenzji oraz wdrożenie funkcji umożliwiających czerpanie z nich korzyści zajęło znacznie więcej czasu. Jednak wyniki badań świadczących o tym, że większe zainteresowanie klientów produktami wywoływane jest w trakcie czytania opinii napisanych przez innych użytkowników, a nie opisów marketingowych, znane były już w 2001 roku [Bickart i Schindler 2001]. Co więcej, w innych badaniach wykazano, że sposób postrzegania przez recenzentów określonych typów produktów i usług może pozwolić na dokładniejsze prognozowanie wyników ich sprzedaży [Dellarocas, Zhang i Awad 2007].

Istnieje kilka strategii, które podmioty zajmujące się sprzedażą produktów i usług mogą obrać w stosunku do fenomenu recenzji konsumenckich. To, jaką strategię dany podmiot może przyjąć, zależy m.in. od tego, jaką rolę w stosunku do danego dobra odgrywa na rynku. Można tu wyróżnić dwie główne role: producenta oraz pośrednika. Producent wytwarza dane dobro i jest z nim silnie powiązany. Negatywna opinia o produkcie lub usłudze może bezpośrednio godzić w jego interes, podczas gdy opinia pozytywna może przynieść mu znaczną korzyść. Natomiast pośrednik jest w stosunku do danego dobra zarówno nabywcą (kupuje je od producenta lub innego pośrednika), jak i sprzedawcą (sprzedaje je kolejnym pośrednikom lub konsumentom). Jego relacja ze sprzedawanymi dobrami nie jest taka silna, jak miało to miejsce w przypadku producenta. Zakładamy, że może zrezygnować z handlu określonym dobrem i, przykładowo, zastąpić jego miejsce w asortymencie dobrem innym.

Analiza recenzji jako sposób na pomiar satysfakcji klienta w celu modyfikacji zachowania sprzedawcy

Podstawowym sposobem interpretacji recenzji konsumenckich przez sprzedawców może być to, że wyrażają one poziom satysfakcji nabywców z produktów i usług. W literaturze znanych jest wiele metod badania satysfakcji klientów, np. systemy skarg i sugestii, wywiady telefoniczne, wywiady osobiste, ankiety elektroniczne czy metoda SERVQUAL [Nieżurawski, Pawłowska i Witkowska 2010]. Podstawowym przytaczanym argumentem, dowodzącym istotności badania satysfakcji klientów jest to, że pozyskanie nowego klienta jest znacznie droższe od kosztu utrzymania klienta stałego [Hill i Alexander 2003].

Monitorowanie recenzji konsumenckich może się odbywać w różny sposób i w różnym zakresie, od czytania przez pracownika recenzji na określonym portalu po systematyczny monitoring całej cyberprzestrzeni przez wyspecjalizowane firmy. Przykładami takich firm działających w Polsce są brand24 (brand24.pl) i Think Kong (www.thinkkong.pl). Odpowiednio prowadzony monitoring opinii wyrażanych w Internecie może nie tylko pozwolić na poznanie aktualnego postrzegania danego dobra przez nabywców, ale również na zaobserwowanie zmian takiego postrzegania w czasie. Może to pozwolić przykładowo na ustalenie tego, czy przyjęta strategia, mająca na celu poprawę takiego sposobu postrzegania dobra, przynosi oczekiwane efekty. Sprzedawcy mogą być również zainteresowani nie tylko sposobem postrzegania produktów i usług oferowanych przez nich, ale również tym, jak w recenzjach konsumenckich oceniane są dobra oferowane przez podmioty konkurencyjne w stosunku do nich.

Znając sposób postrzegania określonego dobra przez nabywców, racjonalnym zachowaniem sprzedawcy jest wykonanie działań, które albo pozwolą zmaksymalizować korzyści wynikające z pozytywnego postrzegania produktu, bądź zminimalizować straty związane z odbiorem negatywnym. Przykładowe akcje, które w odpowiedzi na wyrażane opinie mogą podjąć sprzedawcy to modyfikacja ceny produktu lub usługi, bądź też odpowiednie dostosowanie podejmowanych działań marketingowych do zaistniałej sytuacji. Jest to podobna sytuacja do tej, gdy pewne dobro będzie oceniane w określony sposób w recenzjach eksperckich [Chen i Xie 2005; Godes i in. 2005]. Przykładowo, pozytywne recenzje konsumenckie określonych produktów i usług mogą być wykorzystywane jako element reklamy, zachęcający potencjalnych nabywców do zakupu. Sprzedawcy mogą kupować takie recenzje od portali z opiniami lub sklepów internetowych i zamieszczać je na własnych witrynach [Mudambi i Schuff 2010]. Recenzje konsumenckie mogą być prezentowane potencjalnym nabywcom również za pomocą innych metod, na przykład z wykorzystaniem kana-

łów RSS. W publikacji [Barton 2006] opisano sytuację, gdzie wzbogacenie wiadomości w kanałach RSS o wybrane fragmenty recenzji zwiększyło liczbę kliknięć w hiperłącza w nich zamieszczone o 43%. Patrząc na to zagadnienie z punktu widzenia asymetrii informacji, takie pozytywne recenzje mogą być traktowane jako sygnały pomagające potencjalnym nabywcom rozpoznać jakość produktów [Spence 1973]. Może się to okazać korzystne dla przedsiębiorstw oferujących dobra o wyższej jakości, które w ten sposób mogą bronić się przed wypchnięciem z rynku przez podmioty oferujące gorsze produkty i usługi [Ott, Cardie i Hancock 2012].

Poznanie opinii nabywców może pozwolić również na ustalenie kierunków zmian tak, aby uczynić posiadaną ofertę bardziej odpowiadającą ich potrzebom (funkcja uczenia w monitorowaniu satysfakcji klienta) [Nieżurawski, Pawłowska i Witkowska 2010]. Przykładowo, producenci mogą wykorzystać opinie recenzentów do zmiany produktu lub usługi w taki sposób, aby wyeliminować negatywnie oceniane aspekty. Szczególnie przydatne mogą być w takiej sytuacji konstruktywne opinie konsumentów, zawierające propozycje zmian w danym produkcie lub usłudze [Arora i Srinivasa 2014].

Również pośrednicy mogą wykorzystać recenzje konsumentów do lepszego dopasowania swojego asortymentu do potrzeb nabywców oraz do kształtowania relacji z dostawcami. Pośrednicy muszą podejmować szereg decyzji związanych z określaniem tego, które dobra chcą nabyć od których dostawców, kształtując tym samym swój asortyment, oferowany następnie potencjalnym nabywcom. Identyfikacja sposobu postrzegania dóbr przez konsumentów może pełnić bardzo ważną rolę przy podejmowaniu takich decyzji. Prosty przykładem działania pośrednika jest analiza recenzji w celu wyłonienia produktów najlepiej ocenianych przez recenzentów. Pośrednik może wówczas mieć nadzieję, że jego klienci, usatysfakcjonowani dokonanymi zakupami i dostępnym asortymentem, staną się jego stałymi klientami. W takiej sytuacji, pośrednik po przeanalizowaniu recenzji kupi od dostawców te produkty, które będą w przeanalizowanych przez niego recenzjach najwyżej oceniane. Bardzo istotne może być tu podejmowanie decyzji o współpracy z nowymi dostawcami, np. wchodzącymi na rynek producentami określonych dóbr. Dzięki recenzjom konsumentów pośrednicy mogą szybko zidentyfikować producentów, z którymi warto nawiązać relacje i których produkty warto umieścić w asortymencie. Dzięki odpowiedniej szybkości reagowania na takie sytuacje, możliwe jest przykładowo uzyskanie korzystniejszych warunków współpracy z danym dostawcą.

Oczywiście, istnieje szereg innych strategii, które pośrednicy mogą przyjąć podczas nabywania dóbr od dostawców. Mogą oni analizować recenzje w celu identyfikacji dóbr podobnie postrzega-

nych przez recenzentów, przykładowo w celu zastąpienia w ofercie określonego produktu, handel którym z pewnych przyczyn może być dla pośrednika niekorzystny. Pośrednicy mogą również uwzględniać znajomość specyfiki swoich klientów tak, aby kupować od dostawców dobra, dla których wysoko oceniane były określone aspekty istotne dla grupy docelowej, do której pośrednik chce kierować swoją ofertę.

Udostępnianie możliwości tworzenia i czytania recenzji

Przedsiębiorstwa mogą udostępnić na swoich witrynach internetowych możliwość zamieszczania i oglądania recenzji dóbr, które produkują lub w handlu którymi pośredniczą. Celem wdrożenia takich funkcji może być przykładowo poprawienie funkcjonowania strony z punktu widzenia osiągnięcia przez nią zamierzonego celu. Jako wskaźnik tego, czy wdrożenie systemu recenzji przyniosło spodziewane efekty, często traktuje się tzw. współczynnik konwersji [Gretzel i Yoo 2008], rozumiany jako odsetek osób odwiedzających stronę internetową, które podejmą pożądane przez twórców witryny działanie. Pożądanym działaniem mogą być tu takie akcje użytkownika, jak zakup przedmiotu, założenie konta, pobranie oprogramowania³ itp. Strony publikujące recenzje konsumenckie często mają większe powodzenie od stron nieposiadających takiej funkcji w rozumieniu osiągnięcia wyższego współczynnika konwersji [Gretzel i Yoo 2008].

Dzieje się tak z kilku powodów. Cenioną przez użytkowników funkcją jest możliwość sortowania produktów po ich średnich ocenach czy też wprowadzenie kategorii „najlepiej oceniane produkty” [Barton 2006], dzięki czemu łatwiej jest znaleźć produkty o potencjalnie pożądanej jakości. Dodatkowo, zgodnie z czynnikami warunkującymi decyzje zakupowe konsumentów opisanymi w sekcji 2.5.1, użytkownicy portalu, zastanawiając się nad zakupem określonego produktu czy usługi, w wielu przypadkach częściej mogą się zdecydować na zakup po zapoznaniu się z dostępnymi na witrynie recenzjami konsumenckimi określonego dobra (przy założeniu, że opinie recenzentów o nim będą pozytywne). Duże korzyści wprowadzenia możliwości publikowania recenzji na witrynie wynikają również z faktu, że recenzje są często szukanym typem informacji w Sieci, dzięki czemu istnieje szansa, że na witrynę publikującą takie recenzje trafi wielu użytkowników wyszukiwarek internetowych, którzy inaczej by o jej istnieniu nie wiedzieli [Barton 2006].

Udostępnienie możliwości oceniania produktów i usług może przynieść firmom jeszcze dodatkową korzyść. W przypadku nadawania ocen produktom przez zarejestrowanych użytkowników,

³www.marketingterms.com/dictionary/conversion_rate/. Dostęp: 9 września 2012

takie oceny mogą być analizowane w celu poznania gustów danego użytkownika. W ten sposób możliwe jest generowanie dla niego rekomendacji [Balabanović i Shoham 1997]. Systemy rekomendacyjne dają przedsiębiorstwom możliwość informowania użytkowników o przedmiotach najbardziej odpowiadających ich potrzebom, a więc takich, które ci będą skłonni kupić. Najlepszym przykładem takiego rozwiązania jest sklep amazon.com, którego algorytm ustalania rekomendacji na podstawie zakupionych przez klientów towarów oraz nadawanych przez nich ocen opisano w pracy [Linden, Smith i York 2003]. Dobrze zaprojektowany system rekomendacji pomaga w budowaniu lojalności klientów. Im dłużej bowiem użytkownik jest związany z danym serwisem i im więcej recenzji w nim zamieszcza, tym więcej informacji o jego potrzebach system jest w stanie zebrać, a co za tym idzie, użytkownik otrzymuje ofertę coraz dokładniej dopasowaną do swoich potrzeb. Porównując wzorce oceniania określonego użytkownika z wzorcami innych recenzentów, możliwe jest określanie grup użytkowników o podobnych preferencjach, co pozwala na jeszcze trafniejsze rekomendowanie produktów czy też rekomendowanie recenzji, które mogą oceniać określone dobro z punktu widzenia istotnego dla danego użytkownika [Suresh, Roohi i Eirinaki 2014]. Podsumowując, dzięki udostępnieniu funkcji oceniania produktów i systemom rekomendacji, właściciele serwisów internetowych mogą łatwiej przekształcić sporadycznych gości swoich witryn w stałych klientów [Gaul i Schmidt-Thieme 2001].

Aktywne uczestnictwo w dyskusjach i manipulacje

Podmioty, których produkty lub usługi są recenzowane, mogą oczywiście brać udział w dyskusjach związanych z wyrażaniem opinii na swój temat, gdzie w otwarty sposób mogą się bronić przed krytyką czy oskarżeniami, bądź też przekonywać czytelników do zakupu swoich produktów. Możliwość taka jest szczególnie istotna dla producentów (w tym usługodawców), którzy, jak wspomniano, są szczególnie silnie powiązani z produktami i usługami, które sprzedają. To zwłaszcza w ich interesie leży, aby potencjalni nabywcy nie zniechęcili się do kupna oferowanego przez nich dobra po zapoznaniu się z opiniami zawartymi w recenzjach konsumenckich. Jako przykład portalu umożliwiającego producentom publikowanie odpowiedzi na recenzje warto podać portal gastronauta.pl, na którym funkcjonuje system współpracy z restauracjami. Dzięki niemu, pod każdą opinią możliwe jest dodanie komentarza przez lokal, którego dana opinia dotyczy. W takim komentarzu podmiot może przykładowo ustosunkować się do przedstawionej krytyki. Innym celem uczestnictwa w dyskusji jest podjęcie próby zmiany opinii wobec przedsiębiorstwa oraz jego produktów i usług tych recenzentów, którzy wyrazili negatywne recenzje określonych dóbr (jest

to tzw. naprawcza funkcja w monitorowaniu satysfakcji klienta) [Nieżurawski, Pawłowska i Witkowska 2010]. W celu identyfikacji portali, na których tego typu udział w dyskusji może przynieść firmie największe korzyści, można wykorzystać wspomniane usługi monitorujące reputację firm w Internecie (patrz rysunek 2.4).

Zastosowania Brand24 .

Razem z kontem Brand24, otrzymujesz do dyspozycji potężne narzędzie o wielu zastosowaniach. Głównym przeznaczeniem serwisu jest monitoring marek lub produktów, a także śledzenie tematów lub trendów. Pamiętaj, że pomagamy wychwycić dyskusje kluczowe dla Twojego biznesu niezależnie czy pojawia się w nich nazwa Twojej marki lub produktu. Przykładowo Brand24 pozwala na odnajdywanie osób aktualnie szukających kredytu, skutecznej diety, czy telewizora najlepiej spełniającego ich oczekiwania. Dzięki nam, możesz włączyć się do tych dyskusji i przekonać potencjalnych klientów do skorzystania z Twoich usług lub produktów.

[czytaj poradnik](#)

[pomiń poradnik](#)

[projekt demo](#)

Rysunek 2.4: Materiał marketingowy ze strony brand24. Źródło: zrzut ekranu ze strony brand24.pl/panel/first-steps/ (dostęp po zalogowaniu). Dostęp: 29 stycznia 2013

Analizując sposoby wykorzystywania przez firmy recenzji konsumenckich, nie sposób nie wspomnieć o możliwych nadużyciach, jakich firmy mogą się dopuszczać. W wielu przypadkach, opinie publikowane w Internecie nie podlegają kontroli wiarygodności wygłoszonych w nich stwierdzeń. Zazwyczaj, osoby publikujące takie recenzje mogą również pozostać anonimowe dla zwykłego użytkownika Sieci. Stąd, wiele firm nie opiera się pokusie zamieszczania zmanipulowanych recenzji konsumenckich (oczywiście bez ujawniania prawdziwej tożsamości autora danej recenzji) [Godes i in. 2005]. Działanie takie może odbywać się na różną skalę, od pojedynczej osoby działającej na własną korzyść po zorganizowane grupy, wynajmowane przez firmy w ściśle określonym celu [Liu 2011]. Trudno oszacować rzeczywiste rozmiary tego zjawiska, jednak w pracy [Dellarocas 2006] udowodniono, że w większości sytuacji taka manipulacja ocenami jest naturalną strategią zachowania się sprzedawców. Dwoma najprostszymi sposobami takiej manipulacji jest chwalenie swoich produktów lub marki bądź też wygłaszanie negatywnych opinii o konkurencji [Dellarocas 2006; Liu 2011]. Zjawisko takie w angielskojęzycznej literaturze nosi nazwę *opinion spam* [Liu 2011], zaś fałszywe tożsamości tworzone w celu manipulacji recenzjami noszą nazwę *sock puppets*

(„skarpetka-marionetka”) [Pang i Lee 2008]. Istnieją również bardziej zawaolowane sposoby manipulacji, jak na przykład nagradzanie przez firmy swoich klientów za zamieszczanie pozytywnych opinii o ich produktach i usługach na popularnych forach internetowych [Dellarocas 2006].

Znanych jest wiele udokumentowanych przypadków wspomnianych rodzajów manipulacji. Przykładowo, John Mackey, CEO firmy Whole Foods Market, przez 8 lat używał fałszywej tożsamości, pisząc pochlebne opinie o swojej firmie w serwisie Yahoo! Message Boards [Stone i Richtel 2007]. Ciekawa możliwość poznania praktyk manipulacji recenzjami powstała w wyniku błędu serwisu Amazon.com, który upublicznił prawdziwe tożsamości autorów recenzji. Okazało się, że istniały przypadki, gdy znani pisarze zamieszczali recenzje własnych książek z najwyższą możliwą notą [Harmon 2004]. Innym znanym przykładem podobnego działania była akcja przeprowadzona w trakcie promocji pierwszej płyty piosenkarki Cystiny Aguilery, gdy pracownicy firmy marketingowej zamieszczali wpisy zachwalające płytę na pokojach czatowych i stronach często odwiedzanych przez fanów Britney Spears, piosenkarki o podobnym repertuarze [Mayzlin 2006].

W wyniku takiej działalności firm, nabywcy muszą odpowiednio dopasować swoje zachowanie, uwzględniając we wnioskach wyciąganych z analizy recenzji fakt, że wiele z nich może być zmanipulowanych. Model zaprezentowany w pracy [Dellarocas 2006] pozwala stwierdzić, że w większości przypadków taka działalność firm jest szkodliwa dla nabywców, jednak w długim okresie należy oczekiwać, że dzięki powiększaniu się liczby niezmanipulowanych opinii, firmy nie będą w stanie wpłynąć na zachowania konsumentów za pomocą manipulacji opiniami.

Z uwagi na powszechność manipulacji recenzjami, podejmowane są różnorodne inicjatywy, mające na celu walkę z tym zjawiskiem. Przykładowo, na serwisie Amazon.com wprowadzono specjalne kategorie recenzji, będące z założenia bardziej wiarygodnymi⁴:

- *Amazon Verified Purchase Reviews* – tego typu recenzje mogą być publikowane jedynie przez zalogowanych klientów sklepu, którzy nabyli recenzowane dobro,
- *Amazon Vine* – umożliwia grupie wyselekcjonowanych, najbardziej wiarygodnych klientów dokonywanie recenzji nowych produktów.

Z kolei na portalu *ebay.com* funkcjonuje polityka przeciwko wymuszaniu opinii, mająca na celu wyeliminowanie zjawiska nakłaniania partnerów transakcji do wystawiania konkretnych opinii

⁴na podstawie www.amazon.com/gp/help/customer/display.html?ie=UTF8&nodeId=200791020. Dostęp: 16 września 2012

(przykładowo, aby nie było możliwe wymaganie przez sprzedających na klientach wystawienia lub wycofania konkretnych opinii). Podmioty, które są ofiarami próby wymuszenia opinii, mogą poinformować o tym pracowników portalu, którzy w rezultacie mogą m.in. ograniczyć lub zawiesić konto podmiotów wymuszających ⁵.

Kolejnym sposobem, który może pomóc w walce z nieuczciwymi opiniami, są mechanizmy oceny samych recenzji [Ghose i Ipeirotis 2007]. Na wielu portalach, każda recenzja z osobną również może być oceniana. Przykładowo na portalach `ceneo.pl` oraz `merlin.pl` użytkownicy po zapoznaniu się z recenzją mogą ocenić ją jako przydatną bądź nieprzydatną. Dodatkowo, na portalu `merlin.pl` na tej podstawie tworzony jest ranking recenzentów, których recenzje uznawane są za najbardziej przydatne.

Innym sposobem walki z manipulacją recenzjami jest automatyczne przetwarzanie wielu recenzji w poszukiwaniu podejrzanych wpisów. Stosowane są tym celu różne metody drażenia danych i drażenia tekstu. Analizę trudności z tym związanych i wykorzystywanych w tym celu metod zaprezentowano m.in. w pracy [Liu 2011].

2.5.3 Problem przesytu informacyjnego w kontekście wykorzystania recenzji konsumentów przez podmioty uczestniczące w wymianie rynkowej

Powyżej wskazano, w jaki sposób recenzje konsumentów mogą wpływać na zachowania podmiotów na rynku. Często, dążenie przez określone podmioty do osiągnięcia założonego celu w kontekście recenzji konsumentów wymaga przeprowadzenia analizy wielu takich recenzji. Jak wspomniano we Wprowadzeniu do pracy, w takich sytuacjach może wystąpić problem przesytu informacyjnego.

Podsumowanie takich sytuacji zaprezentowano w tabeli 2.2 na stronie 42. Zidentyfikowane sytuacje można traktować jako punkty wyjściowe do formułowania scenariuszy, dla których istotne byłoby wykorzystanie mechanizmów analizy wydźwięku, umożliwiających automatyczną analizę i podsumowywanie opinii formułowanych w postaci wypowiedzi tekstowych.

2.6 Podsumowanie

W rozdziale omówiono wybrane zagadnienia związane z recenzjami konsumentów publikowanymi w Internecie. Przedstawiono genezę zjawiska oraz omówiono podstawowe pojęcia z nim

⁵pages.ebay.pl/help/policies/feedback-extortion.html

związane (sekcja 2.1), następnie zaproponowano systematykę portali, które umożliwiają publikowanie recenzji, omówiono cechy takich portali i przedstawiono możliwe sposoby wyrażania opinii. Dodatkowo, w sekcji 2.5 przedstawiono wyniki analizy literatury z zakresu wpływu recenzji konsumenckich na zachowanie podmiotów na rynku.

Na podstawie przeprowadzonej analizy można stwierdzić, że recenzje konsumenckie pełnią istotną rolę w działaniach podmiotów funkcjonujących na współczesnym rynku. Recenzje takie można publikować i łatwo odnajdywać na wielu różnych typach portali, przy czym mogą one być wyrażane na różny sposób. Dodatkowo, wpływ takich recenzji na decyzje podejmowane na rynku staje się coraz poważniejszy, co ma wyraz w znacznej liczbie badań poświęconych temu zagadnieniu oraz wynikom sugerującym, że zarówno kupujący, jak i sprzedający przywiązują dużą wagę do recenzji publikowanych w Sieci.

Rola podmiotu na rynku	Sytuacja, w której może wystąpić zjawisko przesytu informacyjnego przy analizowaniu recenzji konsumenckich
Producent	<ul style="list-style-type: none"> • pomiar satysfakcji nabywców na podstawie recenzji konsumenckich, mający umożliwić: <ul style="list-style-type: none"> – odpowiednie dopasowanie działań marketingowych do sposobu postrzegania dobra na rynku, – wprowadzenie modyfikacji produktu lub usługi, mających na celu podniesienie użyteczności dobra dla konsumentów • identyfikacja wypowiedzi lub całych dyskusji, w których wyrażane opinie mogą szkodzić reputacji producenta w celu umożliwienia odpowiedniego zareagowania na pojawiającą się krytykę (np. poprzez wzięcie udziału w dyskusji), bądź na które producent powinien odpowiedzieć z innych powodów
Pośrednik	<ul style="list-style-type: none"> • poznanie opinii recenzentów w celu podjęcia decyzji o kształtowaniu asortymentu oferowanego przez danego pośrednika i relacji z dostawcami; pośrednik może chcieć skonfrontować opinie recenzentów o danym produkcie lub usłudze lub o wszystkich produktach lub usługach określonego producenta z następującymi czynnikami: <ul style="list-style-type: none"> – specyfiką docelowej grupy klientów, – pożądaną strukturą asortymentu (np. eliminowanie substytutów poprzez wykrywanie dóbr ocenianych podobnie przez recenzentów), – stosunkiem ceny do sposobu postrzegania danego dobra przez konsumentów • pomiar satysfakcji klienta w celu odpowiedniego dopasowania działań marketingowych
Konsument	<ul style="list-style-type: none"> • wybór dóbr ocenianych wysoko przez recenzentów (unikanie kupowania produktów i usług ocenianych nisko), • wybór dóbr odpowiadających gustom na podstawie opinii wyrażanych o poszczególnych aspektach dóbr oraz subiektywnych preferencji dotyczących istotności tych aspektów w postrzeganiu produktów i usług

Tabela 2.2: Sytuacje, w których podmioty uczestniczące w wymianie rynkowej mogą zetknąć się ze zjawiskiem przesytu informacyjnego w związku z analizą recenzji konsumenckich. Źródło: opracowanie własne

Rozdział 3

Analiza wydźwięku i metody konstruowania leksykonów sformułowań wyrażających opinie

W poprzednim rozdziale zaprezentowano wyniki badań, które wskazują na znaczącą rolę recenzji konsumenckich w podejmowaniu decyzji przez podmioty uczestniczące w wymianie rynkowej. Jednakże, jak zaznaczono, olbrzymia liczba takich recenzji w Internecie oraz fakt, że mogą być one publikowane w wielu różnych miejscach w Sieci, powodują materializację zjawiska przesytu informacyjnego.

W sekcji 2.4 omówiono różne formaty wyrażania opinii (na przykład za pomocą listy zalet i wad czy też skali punktowej). Zwrócono tam uwagę na fakt, że najbardziej problematycznym formatem z punktu widzenia stopnia trudności automatycznego przetwarzania jest wypowiedź tekstowa. Jednocześnie format ten jest bardzo rozpowszechniony i wiele opinii publikowanych w Internecie wyrażanych jest właśnie w tym formacie. W tym rozdziale zaprezentowana zostanie analiza literatury z zakresu drażenia opinii, czyli automatycznego przetwarzania takich właśnie opinii.

Rozdział ma następującą strukturę. Po krótkim przypomnieniu definicji drażenia opinii, w sekcji 3.2 przeanalizowano pojęcie aspektu i roli aspektów w drażeniu opinii. Następnie przedstawiono główne podejścia do analizy wydźwięku prezentowane w literaturze przedmiotu. W kolejnej sekcji 3.4 skupiono się na zagadnieniu automatycznego pozyskiwania leksykonów sformułowań służących do wyrażania opinii oraz identyfikacji polarności takich sformułowań, co stanowi główny

punkt zainteresowania pracy. Dalej, dokonano przeglądu prac z zakresu drażenia opinii prowadzonych dla języka polskiego. Rozdział kończy się opisem zidentyfikowanych braków w istniejących metodach analizy wydźwięku i krótkim podsumowaniem.

3.1 Drażenie opinii

Drażenie opinii (ang. *opinion mining*), zwane też analizą wydźwięku¹ (ang. *sentiment analysis*) jest dobrze znanym problemem z zakresu przetwarzania języka naturalnego. Z formalnego punktu widzenia, opinię można zdefiniować jako piątkę uporządkowaną, przedstawioną w formule 1.1 na stronie 5. Badania z zakresu analizy wydźwięku zazwyczaj toczą się nad jedynie wybranymi jej elementami. Przykładowo, znaczna część prac skupia się jedynie na określaniu orientacji semantycznej całego dokumentu, bez rozróżniania opinii dotyczących bytów czy ich aspektów opisywanych w tekście np. [Dave, Lawrence i Pennock 2003; Turney 2002]. Drażenie opinii wymaga wykorzystania różnych zasobów i technik w celu rozpoznawania opinii wyrażanych w tekście. Takie zasoby to m.in. zaadnotowane korpusy, na podstawie których można przeprowadzić uczenie maszynowe w celu skonstruowania klasyfikatora, leksykony słów używanych przez konsumentów do wyrażania opinii czy leksykony nazw różnych aspektów recenzowanych produktów i usług. Zasoby takie muszą zostać w jakiś sposób pozyskane lub opracowane.

3.2 Określanie aspektów recenzowanych przedmiotów

Jak wspomniano, w wielu pracach dotyczących drażenia opinii, wprowadzane są pewne ograniczenia w definicji problemu, który usiłuje się rozwiązać. Przykładowo, często usiłuje się jedynie przypisać pojedynczą wartość polarności do całego dokumentu tekstowego. Jednak zgodnie z definicją podaną w sekcji 3.1, drażenie opinii ma na celu określenie szczegółowych opinii, wyrażanych o poszczególnych bytach i ich aspektach wspominanych w tekście. W tej sekcji skupimy się właśnie na pojęciu aspektu. Poniżej przedstawiona zostanie analiza definicji tego pojęcia oraz metod określania, o jakich aspektach danego bytu w tekście wyrażane są opinie.

3.2.1 Aspekt produktu - przegląd definicji

Jednoznaczne zdefiniowanie aspektu bytu jest zadaniem kłopotliwym i w wielu pracach nie wyjaśniono, jakie jest dokładne rozumienie tego pojęcia. W takich przypadkach, autorzy opierają się

¹W prezentowanej pracy słowa *wydźwięk* oraz *opinia* są traktowane jako synonimy.

na jego intuicyjnym rozumieniu. W literaturze anglojęzycznej zamiennie używa się dwóch słów w odniesieniu do aspektu: *feature* oraz *aspect*. W trakcie analizy literatury, zidentyfikowano liczne próby definicji tego pojęcia, gdzie jako aspekt rozumiano ²:

- komponenty (np. obiektyw w aparacie fotograficznym), atrybuty (np. ciężar i wymiary) lub funkcje produktu [Chan i King 2009; Hu i Liu 2004a],
- atrybuty bądź komponenty produktu, do którego recenzenci odnoszą się w swoich opiniach [Liu, Hu i Cheng 2005; Sun i in. 2009],
- kolekcje (zazwyczaj niewielu) termów (czyli pojedynczych słów lub dłuższych sformułowań będących nośnikami znaczenia), mających podobne znaczenie w pewnym kontekście i charakteryzujących określony podtemat pewnej domeny, np. cechy produktu lub atrybuty usługi [Lu i in. 2011; Lu, Zhai i Sundaresan 2009; Wang, Lu i Zhai 2010],
- właściwości, części, cechy części, powiązane koncepty oraz właściwości powiązanych konceptów bytu, gdzie przez powiązane koncepty rozumiane są koncepty istotne dla wrażeń podmiotu wyrażającego opinię o ocenianym bycie [Popescu i Etzioni 2005].

W pracy, jako aspekt bytu (w naszym przypadku dobra, czyli produktu lub usługi) będziemy rozumieć wszystko, co może wpływać na opinię innych podmiotów o tym bycie i co jest z nim bezpośrednio powiązane. Przykładowo, na opinię pewnej osoby o usłudze może wpłynąć jej zły nastrój; czynnik ten nie jest bezpośrednio powiązany z recenzowaną usługą, nie może więc być traktowany jako aspekt tej usługi. Opinia podmiotów o aspekcie może być odmienna od ich opinii o całym bycie lub o innych jego aspektach, jednak musi posiadać pewien wpływ na całościową ocenę bytu przynajmniej dla niektórych podmiotów oceniających dany byt. Dodatkowo, taki wpływ *ceteris paribus* (w tym przypadku, przy założeniu niezmienności oceny innych aspektów) nie może być odwrotny, tj. niemożliwa jest sytuacja, w której pozytywna ocena aspektu pogarsza całościową ocenę podmiotu o danym bycie.

Co istotne, w różnych systemach informatycznych, lista aspektów, które mogą podlegać ocenie przez recenzentów, może być różna. Związane jest z tym pojęcie konceptualizacji, czyli specyfikacji „warstwy pojęciowej tworzącej abstrakcyjny, uproszczony obraz świata” [Abramowicz 2008].

²W części z prac cytowanych poniżej skupiano się na drażeniu opinii wyłącznie o produktach, w związku z czym aspekty były definiowane tylko w odniesieniu do produktów (pomijano w definicjach znaczenie tego słowa w kontekście usług).

Każda osoba może dokonywać innej konceptualizacji otaczającego ją fragmentu rzeczywistości [Abramowicz 2008]. Tym samym, w różnych systemach listy aspektów, które mogą być recenzowane z wykorzystaniem różnych formatów wyrażania opinii i dla których mogą być przeprowadzane wnioski, mogą być inne.

Do każdego aspektu przypisana jest kolekcja termów, dla których dany aspekt jest desygnatem. Poszczególne terminy tworzące opisaną kolekcję będziemy określać jako **nazwy aspektu**. Opinie o aspektach w tekście mogą być wyrażane albo wprost, jeśli w tekście pojawia się słowo będące nazwą danego aspektu lub też pośrednio, jeśli nazwa się nie pojawia, jednak możliwe jest wywnioskowanie, że dany aspekt podlega ocenie na podstawie innych słów pojawiających się w tekście [Liu, Hu i Cheng 2005; Popescu i Etzioni 2005]. Przykładowo, słowa *ciężki* i *duży* nie są nazwami aspektów, lecz mogą być ich ocenami (w tym przypadku przykładowo takich aspektów jak *waga* i *rozmiar*). Jeśli słowa te wystąpią w tekście, aspekty, do których się odnoszą, mogą być łatwo rozpoznane przez czytelnika.

Dla danego podmiotu, różne aspekty mają różny wpływ na całościową ocenę bytu. Inaczej mówiąc, niektóre aspekty mogą być dla tego podmiotu ważniejsze niż inne. Przykładowo, dla wielu konsumentów, w ocenie telefonu komórkowego ważniejsza będzie wytrzymałość baterii niż fakt wyposażenia go w czujnik wilgoci [Yu i in. 2011], jednak nie należy oczekiwać, że zależność ta będzie prawdziwa dla wszystkich bez wyjątku konsumentów.

Aspekty mogą pozostawać względem siebie w relacji hierarchicznej. Przykładowo, produkt *laptop* może posiadać aspekt *bateria*, który z kolei może być uszczegóławiany przez takie aspekty, jak *rozmiar baterii*, *waga baterii*, *pojemność baterii* itd. [Liu, Hu i Cheng 2005]. Z tego powodu może okazać się, że system podsumowujący opinie pokazuje użytkownikowi oceny na zbyt mało szczegółowym poziomie. Przykładowo, często ocenianym w recenzjach restauracji aspektem jest *jedzenie*, jednak prosta polarność oceny tego aspektu nie mówi nic o tym, z jakiego punktu widzenia został on oceniony. Przykładowo, czy brano był pod uwagę jedynie smak dania, temperatura podania, czy może także jego wpływ na zdrowie konsumentów itp. [Lu, Zhai i Sundaresan 2009].

Kolekcję dóbr, które mogą być opisane za pomocą tej samej kolekcji aspektów, będziemy nazywać **kategorią** dóbr. Takie kategorie to np. restauracje czy telefony komórkowe. Również tutaj możliwe byłoby wskazanie istnienia hierarchii kategorii bytów, jednak dla potrzeb pracy nie jest konieczne głębsze analizowanie tego zagadnienia.

3.2.2 Leksykony nazw aspektów

Aby przeprowadzić analizę wydźwięku z uwzględnieniem opinii o poszczególnych aspektach bytów, konieczne jest posiadanie zasobu wiedzy o tym, jakie aspekty dany byt może posiadać oraz jakie są ich nazwy.

Aspekty bytów oraz nazwy, jakie można tym aspektom przypisać, dla danej kategorii mogą być określone przez eksperta na podstawie znajomości domeny lub przeprowadzonej analizy korpusu tekstów (jak np. w pracach [Lula i Wójcik 2011; Taboada i in. 2011]). Należy jednak pamiętać, że dla każdej kategorii taka lista może być inna (np. inne aspekty będą miały kategorie telefonów komórkowych i restauracji), co znacząco zwiększa nakład pracy eksperta.

Identyfikację aspektów i ich nazw można również przeprowadzać automatycznie na podstawie analizy korpusu recenzji. Zadanie to jest podobne do ekstrakcji słów kluczowych specyficznych dla danej domeny. Problem ten porusza m.in. praca [Frank i in. 1999]. Podstawową informacją, wykorzystywaną przez takie metody, jest informacja o częściach mowy dla poszczególnych słów w tekście ze względu na fakt, że nazwami aspektów zazwyczaj są rzeczowniki [Hu i Liu 2004a]. Słowa wykorzystywane w recenzjach i będące rzeczownikami są więc naturalnymi kandydatami na nazwy aspektów.

Wybór rzeczowników faktycznie będących nazwami aspektów i odfiltrowywanie słów niewłaściwych często odbywa się na drodze statystycznej analizy korpusu. W pracy [Scaffidi i in. 2007] analizuje się częstość występowania poszczególnych rzeczowników w korpusie recenzji bytów z danej kategorii oraz w drugim korpusie, zawierającym dokumenty na różnorodne tematy (tzw. korpus generyczny). Wyrazy, które zdecydowanie częściej pojawiają się w korpusie recenzji niż w korpusie generycznym, są prawdopodobnie istotne dla danej kategorii produktów, stąd istnieje prawdopodobieństwo, że są nazwami aspektów.

Inny, heurystyczny rodzaj podejścia do identyfikacji nazw aspektów zaproponowano w pracach [Hu i Liu 2004a,b]. Heurystyka działała w ten sposób, że jako nazwy aspektów identyfikowano te rzeczowniki, w pobliżu których pojawiały się znane systemowi sformułowania służące do wyrażania opinii. Przykładowo, jeśli w recenzji aparatu fotograficznego pojawiło się sformułowanie *zdjęcia są fantastyczne*, a wyraz *fantastyczne* jest znanym sformułowaniem służącym do wyrażania opinii, możliwe jest automatyczne wywnioskowanie, że *zdjęcia* są nazwą pewnego aspektu aparatu.

Warto wspomnieć również o podejściach wykorzystujących gotową, predefiniowaną taksonomię aspektów konkretnego bytu, które następnie usiłują automatycznie poszerzyć ją o inne nazwy

znanych już aspektów, z wykorzystaniem różnego rodzaju zasobów leksykalnych [Carenini, Ng i Zwart 2005; Liu, Hu i Cheng 2005].

Podsumowując, analiza tekstu w celu identyfikacji aspektów, których dany tekst dotyczy, wymaga wykorzystania odpowiednich zasobów dotyczących aspektów i ich nazw. Zasoby te mogą być przygotowane manualnie, jednak istnieją również prace z zakresu automatyzacji tego procesu.

3.3 Podejścia do analizy wydźwięku

Istnieją dwa główne podejścia do automatycznego określania opinii zawartej w tekście [Taboada i in. 2011]:

1. określanie opinii poprzez statystyczną klasyfikację tekstów i przypisywanie ich do klas reprezentujących różne wartości orientacji semantycznej,
2. określenie wydźwięku za pomocą reguł i leksykonów sformułowań służących do wyrażania opinii wraz z przypisanymi do nich polarnościami.

W tej sekcji zaprezentowane zostaną wyniki analizy literatury dotyczącej obu tych podejść.

3.3.1 Statystyczna klasyfikacja tekstów i nadzorowane uczenie maszynowe

Klasycznym podejściem do rozpoznawania wydźwięku dla opinii wyrażanych w tekście jest statystyczna klasyfikacja tekstów z wykorzystaniem uczenia maszynowego [Gamon 2004]. Program komputerowy jest określany jako uczący się na podstawie doświadczenia E w odniesieniu do pewnej klasy zadań T i miary wydajności P , jeśli jego wydajność wykonywania zadań z klasy T (mierzona za pomocą P) wzrasta wraz z doświadczeniem E [Mitchell i in. 1997]. Intuicyjnie, uczenie maszynowe można wyjaśnić jako dziedzinę nauki, której celem jest umożliwienie komputerom uczenia innymi sposobami niż na drodze bezpośredniego zaprogramowania ich dokładnego działania³.

Zastosowaniem uczenia maszynowego, istotnym z punktu widzenia prezentowanej pracy, jest klasyfikacja tekstu. W problemie tym dysponujemy dwoma głównymi elementami: reprezentacją tekstu $d \in \mathbb{X}$ (gdzie \mathbb{X} to tzw. przestrzeń dokumentów) i ustaloną kolekcją klas $\mathbb{C} = \{c_1, c_2, \dots, c_n\}$,

³Przedstawioną definicję w licznych materiałach dydaktycznych dostępnych w Internecie przypisuje się Arthurowi Samuelowi (miał ją sformułować w roku 1959), jednak autor nie zdołał odnaleźć wiarygodnej publikacji naukowej potwierdzającej ten fakt.

do których dokumenty mają być przypisywane [Manning, Raghavan i Schütze 2008]. Przykładowym zastosowaniem uczenia maszynowego w klasyfikacji tekstu jest rozpoznawanie spamu [Zhang, Su i Wang 2007], gdy każdy dokument (np. wiadomość e-mail) może trafić do jednej z dwóch klas, gdzie jedna z nich odpowiada wiadomościom spamowym, a druga wiadomościom nie będącym spamem. W przypadku zadania analizy wydźwięku, w którym, przykładowo, do każdego tekstu chcemy przypisać polarność pozytywną, negatywną i neutralną, klasy będą odpowiadały właśnie tym polarnościami [Sharma i Dey 2012].

Nadzorowane uczenie maszynowe i problem klasyfikacji tekstu

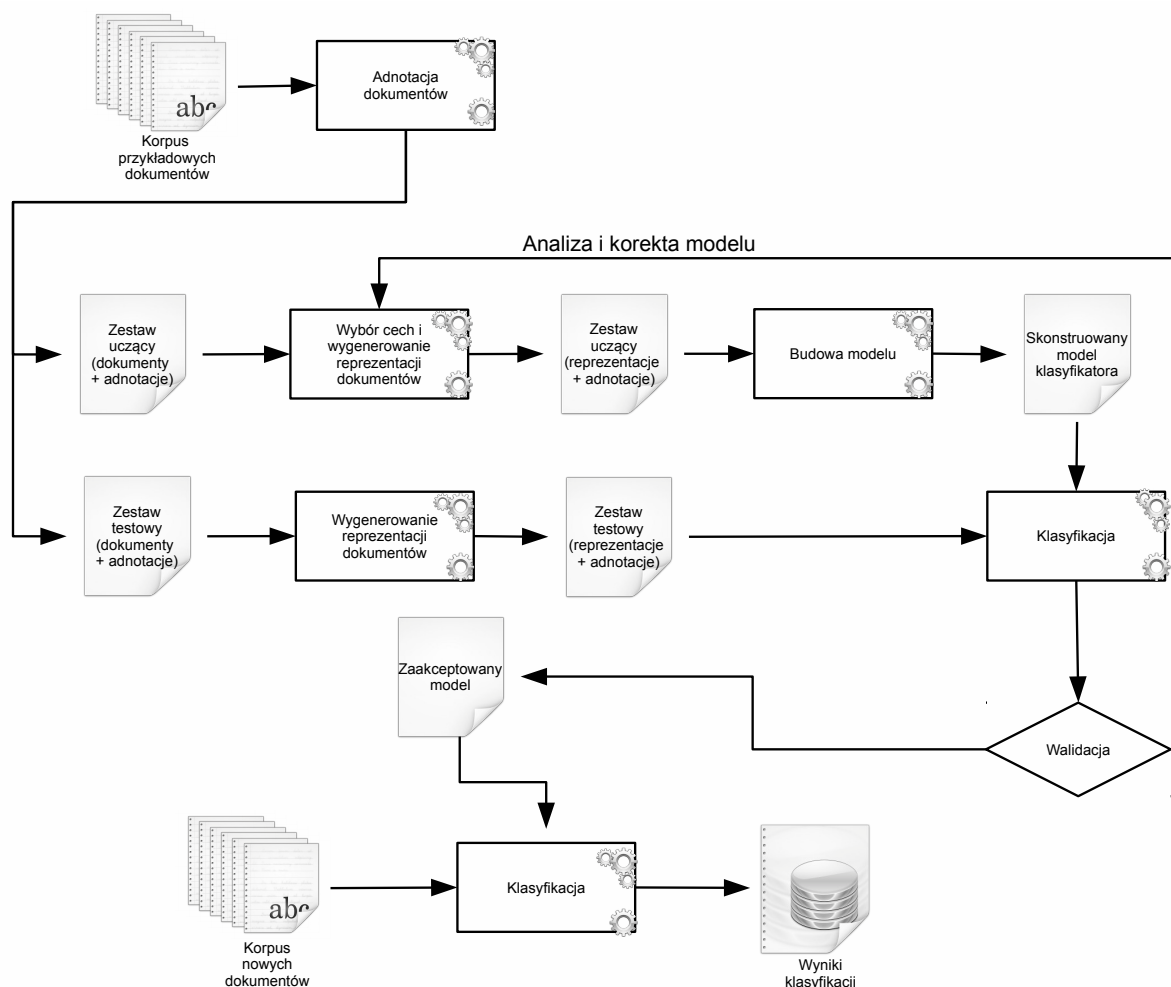
Zgodnie z definicją podaną powyżej, systemy mogą uczyć się na podstawie doświadczenia, dzięki czemu możliwy jest wzrost ich wydajności. W praktyce, doświadczenie przejawia się w dostępności przykładowych danych, dla których zdefiniowano oczekiwane rezultaty działania systemu, tj. klasy, do których te przykłady mają być przypisane. Proces określania oczekiwanych rezultatów pracy systemu dla przykładowych danych będziemy nazywać procesem adnotacji. Proces ten polega na ręcznym przypisaniu przykładowych tekstów do możliwych kategorii np. przez ekspertów. Inną możliwością ręcznej adnotacji jest wykorzystanie usług takich, jak *Amazon Mechanical Turk*⁴, która udostępnia funkcję *crowdsourcingu* prostych zadań wymagających ludzkiej inteligencji [Taboada i in. 2011].

Przypisanie przykładów do klas może się również odbyć automatycznie na podstawie dodatkowych informacji, takich jak np. ocena wyrażona w łatwy do przetworzenia sposób. W pracy [Pang, Lee i Vaithyanathan 2002] wykorzystywano ocenę w skali od 1 do 5, którą recenzenci przypisywali do produktów podczas wyrażania opinii (każda recenzja składała się z wypowiedzi tekstowej i oceny punktowej). Przykładowo, gdy do danej recenzji przypisana była wysoka ocena, adnotowano dany dokument jako należący do klasy odpowiadającej polarności pozytywnej.

Kolejnym kluczowym krokiem w uczeniu maszynowym jest transformacja dokumentów do reprezentacji odpowiedniej do nauki algorytmu wybranego do klasyfikacji [Chaovalit i Zhou 2005; Joachims 1998]. Przykładem takiej reprezentacji jest wektor, gdzie kolejne elementy odpowiadają częstości występowania określonych słów [Sebastiani 2002]. Poszczególne elementy takich wektorów będziemy nazywać cechami, a zatem wektory reprezentujące teksty, będziemy określać mianem wektorów cech. W literaturze znajdują się wyniki eksperymentów dla wektorów reprezentujących różne cechy klasyfikowanych dokumentów. Kolejne elementy wektorów mogły

⁴<https://www.mturk.com>

odpowiadać takim cechom, jak przymiotniki pochodzące z ustalonej kolekcji, n -gramy różnego rzędu⁵, n -gramy wzbogacone o przypisane im części mowy itp. Wartości kolejnych elementów wektorów mogą być określane przykładowo na podstawie miary $TF - IDF$ dla wspomnianych elementów, ich pozycji w tekście itp. [Pang i Lee 2008].



Rysunek 3.1: Schemat klasyfikacji dokumentów tekstowych z wykorzystaniem nadzorowanego uczenia maszynowego. Źródło: opracowanie własne na podstawie diagramu opublikowanego w artykule [Giménez 2010]

Proces uczenia maszynowego, oceny jego skuteczności i zastosowania jego wyników zaprezentowano na diagramie na rysunku 3.1. Istotnym krokiem procesu jest podział posiadanych

⁵N-gramy są to sekwencje wspólnie występujących w tekście (tj. przylegających, występujących w tekście dokładnie w określonej kolejności) określonych jednostek leksykalnych, np. tokenów, gdzie n oznacza długość sekwencji (np. liczbę tokenów)

przykładów na zestaw uczący i testowy, gdzie pierwszy z nich jest wykorzystywany na etapie budowy modelu, natomiast na podstawie drugiego następuje ewaluacja działania systemu. Na podstawie zestawu uczącego i wektorów cech przypisanych dla przykładów uczących, z wykorzystaniem odpowiednich metod następuje nauka modelu służącego do klasyfikacji. Nauka ma na celu skonstruowanie funkcji γ , odwzorowującej wektory cech \mathbb{X} na klasy \mathbb{C} [Manning, Raghavan i Schütze 2008]:

$$\gamma : \mathbb{X} \rightarrow \mathbb{C} \quad (3.1)$$

Posiadając model (klasyfikator), kolejnym krokiem jest przetestowanie poprawności jego działania poprzez klasyfikację przykładów z zestawu testowego i określenie zgodności wyniku z klasami, do których poszczególne przykłady zostały przypisane na etapie adnotacji. W przypadku braku satysfakcjonujących wyników, cały proces można powtórzyć, wybierając inny zestaw cech do wektora lub wybierając inną metodę uczenia maszynowego. Gdy system będzie działał z satysfakcjonującą poprawnością, może być wykorzystany do klasyfikacji nowych tekstów. Możliwe jest również przygotowanie dodatkowego zestawu testowego, za pomocą którego określone będą ostateczne wyniki działania systemu (w tym przypadku istotne jest, aby ten dodatkowy zestaw nie był wcześniej wykorzystywany w przeprowadzanych eksperymentach).

Powyżej wspomniano o tym, że istnieją odpowiednie metody, za pomocą których odbywa się uczenie systemu oraz klasyfikacja. Dla potrzeb analizy wydźwięku, w literaturze przetestowano m.in. naiwny klasyfikator Bayesa, klasyfikator oparty o maszyny wektorów wspierających oraz wykorzystujący modelowanie języka. W celu dokładniejszego wyjaśnienia sposobu działania takich podejść, warto omówić przebieg klasyfikacji dokumentów z wykorzystaniem tych trzech metod.

Naiwny klasyfikator Bayesa

Często stosowaną metodą uczenia maszynowego są naiwne klasyfikatory Bayesa. Jest to przykład probabilistycznego podejścia do klasyfikacji, gdzie dla każdego klasyfikowanego obiektu określa się prawdopodobieństwo jego przynależności do poszczególnych klas. Obiekt jest przypisywany do tej klasy, dla której prawdopodobieństwo jest największe.

Podejście to można zapisać w następujący sposób [Manning, Raghavan i Schütze 2008; Pang i Lee 2008]:

$$c_{map} = \arg \max_{c \in \mathbb{C}} P(c|d) \quad (3.2)$$

Przeanalizujmy prosty przypadek, w którym wektor cech składa się z wartości binarnych $\langle t_1, t_2, \dots, t_n \rangle$, gdzie każda z nich oznacza, czy dane sformułowanie (token) z kontrolowanego słownika występuje w dokumencie, którego reprezentacją jest wektor. W przypadku zastosowania tej metody dla potrzeb analizy wydźwięku, jako tokeny często traktuje się przymiotniki czy sformułowania wykorzystywane do wyrażania opinii [Pang, Lee i Vaithyanathan 2002]. Prawdopodobieństwo $P(c|d)$ jest tu zdefiniowane zgodnie z twierdzeniem Bayesa jako [Manning, Raghavan i Schütze 2008]:

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n} P(t_k|c) \quad (3.3)$$

gdzie:

- $P(c)$ to tzw. prawdopodobieństwo *a priori* przypisania dokumentu do klasy c , obliczane jako estymator największej wiarygodności, tj. stosunek liczby dokumentów należących do danej klasy do wszystkich dokumentów w kolekcji uczącej,
- $P(t_k|c)$ to prawdopodobieństwo wystąpienia słowa t_k w dokumencie należącym do klasy c , również obliczane z wykorzystaniem estymatora największej wiarygodności.

Nauka polega w tej metodzie na obliczaniu estymatorów wspomnianych powyżej prawdopodobieństw. Natomiast klasyfikacja dokumentów, jak już wspomniano, polega na wyborze tej klasy, dla której uzyskano największe prawdopodobieństwo zgodnie z formułą 3.3.

Należy zwrócić uwagę na fakt, że równanie 3.3 wymaga, aby poszczególne prawdopodobieństwa warunkowe były od siebie statystycznie niezależne. Założenie to w wielu przypadkach nie jest spełnione. Pomimo tego, metoda w wielu zastosowaniach pozwala uzyskać dobre rezultaty i w pracy [Smeureanu i Bucur 2012] pozwoliła na osiągnięcie około 80% skuteczności rozpoznawania polarności poszczególnych zdań wchodzących w skład recenzji konsumenckich.

Maszyny wektorów wspierających

Jak wspomniano, w omawianych podejściach każdy dokument jest reprezentowany w postaci wektora o określonej liczbie cech. Każdy wektor odpowiada określonemu punktowi w wielowymiarowej przestrzeni cech (przestrzeń posiada tyle wymiarów, ile cech brane jest pod uwagę w reprezentacji dokumentów) [Joachims 1998]. W związku z tym, problem klasyfikacji można sprowadzić do zdefiniowania obszarów w tej przestrzeni odpowiadających poszczególnym klasom. Klasyfikacja

odbywałyby się w ten sposób, że dany obiekt byłby przypisywany do klasy na podstawie tego, w którym obszarze przestrzeni znajduje się odpowiadający mu wektor [Manning, Raghavan i Schütze 2008; Sharma i Dey 2013].

Jedną z metod uczenia maszynowego, która działa właśnie na zasadzie określenia obszarów przestrzeni i klasyfikacji na podstawie położenia wektora, nosi nazwę maszyny wektorów wspierających (ang. *Support Vector Machine*, w skrócie SVM). Podejście to wykorzystuje założenie, iż dla danego zestawu uczącego, w którym występują punkty odpowiadające osobnym klasom, funkcje dyskryminujące powinny przebiegać pośrodku obszarów pomiędzy zbiorami punktów należącymi do poszczególnych klas. Nauka polega więc na wyznaczeniu takich właśnie funkcji dyskryminujących. Rozwiązanie omawianego zagadnienia jest problemem z zakresu programowania matematycznego i wykracza poza zakres prezentowanej pracy. Warto jedynie nadmienić, że istnieje wiele implementacji komputerowych możliwych do wykorzystania w tym celu (często stosowanym programem jest SVM^{light} [Joachims 1999]). Dokładność klasyfikacji wydźwięku w eksperymentach przeprowadzonych w pracy [Pang, Lee i Vaithyanathan 2002] w najlepszych przypadkach sięgała 83%.

Klasyfikacja z wykorzystaniem modelowania języka naturalnego

Kolejnym podejściem związanym z uczeniem maszynowym jest modelowanie językowe. W tym podejściu, na podstawie przykładowych dokumentów z klasy polarności pozytywnej i negatywnej, generowane są tzw. modele językowe, definiowane jako rozkład prawdopodobieństwa na n -gramach (patrz definicja na str. 50). Określane jest tu prawdopodobieństwo pojawienia się poszczególnych n -gramów w wypowiedziach przypisanych do określonej klasy [Chen i Goodman 1996; Hu i in. 2007].

Omawiane podejście w kontekście drażenia opinii zostało przebadane w artykułach [Bonev, Ramírez-Sánchez i Rojas 2012; Cui, Mittal i Datar 2006; Hu i in. 2007]. W pracy [Hu i in. 2007] skonstruowano modele języków wykorzystywanych przy formułowaniu pozytywnych i negatywnych opinii na n -gramach słów rzędu drugiego (czyli ciągach dwóch słów). Eksperymenty nad n -gramami wyższego rzędu (do $n=6$) zaprezentowano z kolei w artykule [Cui, Mittal i Datar 2006]. Należy jednak zaznaczyć, że dla dłuższych n -gramów coraz istotniejszy staje się rozmiar posiadanego korpusu służącego do nauki, gdyż przy krótkim korpusie prawdopodobieństwo zaobserwowania dłuższych n -gramów może się znacznie zmniejszać.

Jako wynik budowy modelu konstruowane były dwa wzorcowe modele języka, gdzie jeden z

nich dotyczył języka wykorzystywanego w recenzjach o wydźwięku negatywnym, a drugi pozytywnym. Posiadając dwa modele języków, podobieństwo pomiędzy nimi może zostać określone przykładowo z wykorzystaniem dywergencji Kullbacka-Leiblera⁶ [Hu i in. 2007]. Wykorzystując miarę podobieństwa określoną w ten sposób, nowe dokumenty klasyfikowane są na podstawie tego, do którego ze wzorcowych rozkładów jest bardziej podobny rozkład skonstruowany na podstawie danego artykułu. W eksperymentach przeprowadzonych w pracy [Hu i in. 2007], podejście takie pozwoliło na uzyskanie lepszej skuteczności niż w przypadku klasyfikacji z wykorzystaniem maszyn wektorów wspierających.

3.3.2 Analiza wydźwięku z wykorzystaniem reguł i leksykonów sformułowań służących do wyrażania opinii

Innym podejściem do analizy wydźwięku jest wykorzystanie przetwarzania regułowego i leksykonów sformułowań wyrażających opinie do określenia polarności tekstu. Podejście to wymaga posiadania leksykonu sformułowań, za pomocą których recenzenci mogą wyrażać opinie, wraz z przypisaną informacją o tym, jaką opinię (pozytywną, negatywną czy neutralną) dane sformułowanie określa [Ding, Liu i Yu 2008; Taboada i in. 2011]. Sposoby przygotowywania takich leksykonów zostały opisane w sekcji 3.4.

Posiadając leksykony sformułowań o znanych polarnościach, możliwe jest przetwarzanie recenzji z wykorzystaniem zdefiniowanych reguł i heurystyk. Przykładem reguły może być np. wzorzec ekstrakcji, który wyszukuje odpowiednie sekwencje, zawierające nazwy aspektów oraz sformułowania z leksykonu sformułowań służących do wyrażania opinii, jak to miało miejsce w pracy [Popescu i Etzioni 2005]. Prosta heurystyką może być z kolei określanie polarności całego dokumentu poprzez zliczenie liczby słów z leksykonów o polarności pozytywnej i negatywnej, które w tym dokumencie wystąpiły i przypisanie polarności na podstawie tego, która liczba jest większa [Dave, Lawrence i Pennock 2003; Hu i Liu 2004a].

Prosty przykład potoku przetwarzającego teksty w opisany sposób w celu przeprowadzenia analizy wydźwięku na poziomie aspektów może wyglądać następująco [Liu 2011]:

1. zaznacz wszystkie wystąpienia nazw aspektów w tekście oraz znane sformułowania służące do wyrażania opinii,

⁶Pomimo tego, że miara ta nie spełnia definicji metryki, często w praktyce jest interpretowana jako odległość pomiędzy dwoma rozkładami.

2. przypisz wszystkim słowom o znanej pozytywnej polarności wartość $+1$, zaś tym o polarności negatywnej -1 ,
3. przetwórz wyrazy takie jak negacje (*nie, żadne, wcale* itp.) oraz czasowniki modalne (*mógłby, powinien*), które często zmieniają orientację opinii w znajdujących się tuż za nimi słowach [Thet, Na i Khoo 2010; Wilson, Wiebe i Hoffmann 2005], np. zignoruj wszystkie słowa znajdujące się w ich bezpośrednim otoczeniu lub odwróć ich polarność,
4. zagreguj opinię o aspektach, co może się odbyć w następujący sposób:
 - połącz w pary nazwy aspektów i sformułowania służące do wyrażania opinii, np. na podstawie odległości w tekście (liczonej w wyrazach) pomiędzy nimi [Chan i King 2009; Liu i in. 2014],
 - dla każdego aspektu, zsumuj polarności przypisane do odpowiadających mu sformułowań służących do wyrażania opinii,
 - ustal opinię o aspekcie na podstawie tego, czy otrzymana suma polarności jest pozytywna czy negatywna [Dave, Lawrence i Pennock 2003].

Opisane kroki są jedynie bardzo podstawowym podejściem do rozwiązania analizowanego problemu. Nie uwzględniono tu m.in. faktu, że poszczególne sformułowania służące do wyrażania opinii mogą mieć różną polarność w zależności od aspektu, który opisują. W literaturze można napotkać wiele dalszych reguł i heurystyk, które przypisują orientację na podstawie sekwencji i typów słów, które pojawiają się w zdaniu. Przykładowo, dodatkową informację mogą nieść przysłówki, które można traktować jako dodatkowy wymiar reprezentujący intensywność wyrażonej opinii [Turney i Littman 2003] (np. *bardzo, niewiarygodnie* itp.). W prostym podejściu, przysłówki pojawiające się w odległości trzech wyrazów od sformułowań służących do wyrażania opinii mogą być do nich przypisywane. W takim przypadku, w końcowym obliczaniu opinii takie słowa mogą mieć przypisane większe znaczenie [Chen, Qi i Wang 2012; Dalal i Zaveri 2014].

3.4 Automatyczne pozyskiwanie leksykonów sformułowań służących do wyrażania opinii

Kluczowym zasobem wykorzystywanym w wielu pracach z zakresu drażenia opinii jest leksykon sformułowań służących do wyrażania opinii. W podejściu wykorzystywanym do uczenia maszy-

nowego, poszczególne słowa z takiego leksykonu mogą odpowiadać cechom w wektorze cech, służącym do opisu klasyfikowanego tekstu [Pang i Lee 2008]. Natomiast w podejściu regułowym, opisanym w podsekcji 3.3.2, sformułowania z takiego leksykonu były brane pod uwagę podczas formułowania reguł i heurystyk [Liu 2011].

Sformułowania w leksykonie mogą mieć polarność pozytywną (np. *świetny, cudowny*) lub negatywną (*kiepski, beznadziejny*) w stosunku do opisywanego przedmiotu lub aspektu przedmiotu. Przynależność do kolekcji słów pozytywnych lub negatywnych może być rozumiana jako zero-jedynkowa, gdzie słowa należą tylko do jednej z tych kolekcji (lub do żadnej z nich, jeśli słowo nie wyraża opinii), mogą być przypisywane do nich różne poziomy intensywności [Chaovalit i Zhou 2005], bądź też mogą być rozumiane w kategoriach rozmytych [Andreevskaia i Bergler 2006]. Możliwe jest również uszczegółowienie tych kolekcji poprzez ich podział na konkretne typy pozytywnych i negatywnych emocji [Grefenstette i in. 2006].

3.4.1 Problem względności polarności sformułowań

Istnieją słowa, które wyrażają subiektywne odczucia osoby wydającej opinię i zachowują swoje znaczenie niezależnie od przedmiotu, który opisują [Kanayama i Nasukawa 2006; Takamura, Inui i Okumura 2006]. Przykładami słów tego typu są wyrazy opisujące sześć podstawowych emocji - gniewu, zniesmaczenia, strachu, radości, smutku i zaskoczenia [Pang i Lee 2008]. We wczesnych badaniach nad drażnieniem opinii, przykładowo w pracach [Nasukawa i Yi 2003; Subasic i Huettnner 2001], leksykony takie bywały konstruowane ręcznie. Istnieją również publicznie dostępne zasoby, które w założeniu twórców mają dawać możliwość wykorzystania ich w charakterze generycznych leksykonów takich sformułowań. Przykładowo, dla języka angielskiego takim zasobem jest SentiWordNet [Baccianella, Esuli i Sebastiani 2010]. W badaniach dla języka polskiego, opisanych w pracach [Buczynski i Wawer 2008a,b], wykorzystano słownik przygotowany przez firmę zajmującą się monitoringiem reputacji w Internecie o nazwie Zetema. Podejmowano również próby skonstruowania metod uczących się polarności sformułowań wyrażających opinie na podstawie analizy słowników czy innych zasobów językowych (prace [Godbole, Srinivasaiiah i Skiena 2007; Hu i Liu 2004a] mogą służyć tu za przykłady).

Należy jednak zwrócić wagę na fakt, że, jak wspomniano we wprowadzeniu do pracy, polarność związana z danymi sformułowaniami silnie zależy od kontekstu, w którym są one użyte, czyli od bytu czy też aspektu, do którego się odnoszą [Liu i in. 2014]. Często również, sformułowania, które same z siebie nie mają żadnej orientacji semantycznej, po połączeniu we frazy z innymi sło-

wami uzyskują pewną orientację, zależną od bytu i aspektu, do którego się odnoszą [Takamura, Inui i Okumura 2006]. Dodatkowo, sformułowania takie nie muszą odnosić się do żadnej emocji. Mogą one opisywać pewien fakt, a dopiero czytelnik, wykorzystując swą wiedzę, nadaje danemu słowu pozytywną lub negatywną interpretację [Kanayama i Nasukawa 2006]. Przykładowo, słowo „nieprzewidywalny” zmienia swoją orientację z pozytywnej w przypadku opinii o filmach („nieprzewidywalna fabuła”) na negatywną w domenie samochodów („nieprzewidywalne zachowanie samochodu na zakręcie”). Podobnie, sformułowanie „straszna scena” ma pozytywny wydźwięk w przypadku recenzji horroru, pomimo tego, że samo słowo „straszny” w wielu innych kontekstach miałoby wydźwięk przeciwny [Chaovalit i Zhou 2005]. Problem można nakreślić jeszcze bardziej szczegółowo, gdyż te same słowa, używane w tekstach w tej samej domenie, mogą mieć przeciwną orientację w zależności od aspektu, do którego się odnoszą (np. „długi czas życia baterii” i „długi czas ustawiania ostrości” w recenzji aparatu fotograficznego) [Liu 2011; Lu i in. 2011].

W związku z opisywaną sytuacją, w leksykonie muszą znajdować się nie tylko sformułowania i przypisane do nich polarność, ale też informacje o tym, w stosunku do jakiego bytu i jakiego aspektu dane sformułowanie ma określoną polarność. Prowadzono badania mające na celu dostosowywanie istniejących leksykonów sformułowań wyrażających opinie przy zmianie domeny (przykładowo w pracy [Choi i Cardie 2009]), jednak zazwyczaj problem ten rozwiązuje się poprzez przygotowanie nowego, specyficznego dla danej domeny leksykonu.

Jak już wspomniano, leksykony sformułowań wykorzystywanych do wyrażania opinii mogą być tworzone manualnie. Istnieje jednak wiele publikacji dotyczących metod automatycznego lub półautomatycznego generowania słownika oraz przypisywania polarności do sformułowań w nim zawartych. Można tu wyodrębnić dwa zadania:

- identyfikacja sformułowań służących do wyrażania opinii,
- przypisywanie polarności do zidentyfikowanych sformułowań.

Oba te zadania mogą być realizowane jednocześnie, jak to miało miejsce np. w artykule [Hatzivassiloglou i McKeown 1997], jednak mogą one również być rozłączne. Dodatkowo, można rozróżnić dwa podejścia do realizacji tych zadań: [Andreevskaia i Bergler 2006; Liu 2011]:

- metody słownikowe, wykorzystujące zazwyczaj zasoby takie jak WordNet i analizujące znajdujące się w nich definicje wyrazów i relacje między nimi,
- metody wykorzystujące analizę korpusu, polegające na analizie dużych korpusów tekstów z wykorzystaniem np. wzorców syntaktycznych, analizy współwystępowania wyrazów itp.

W dalszej części sekcji zaprezentowany zostanie przegląd prac dotyczących tych zagadnień.

3.4.2 Identyfikacja sformułowań służących do wyrażania opinii

Etap identyfikacji sformułowań służących do wyrażania opinii ma na celu pozyskanie takich sformułowań do leksykonu. Abstrahujemy tu od przypisywania do tych sformułowań polarności, co można traktować jako osobny krok w procesie nauki do analizy wydźwięku [Arora i Srinivasa 2014].

Podstawowa możliwość identyfikacji sformułowań służących do wyrażania opinii wynika z faktu, że często sformułowania takie pojawiają się w recenzjach w określonych sekwencjach z innymi wyrazami lub z wyrazami będącymi określonymi częściami mowy. Sekwencje takie będziemy nazywali wzorcami (ang. *patterns*) [Grefenstette i in. 2006]. W związku z tym, analiza korpusu i wyszukiwanie określonych sekwencji może pozwolić na identyfikację sformułowań służących do wyrażania opinii. Przykładowo, w pracy [Wawer 2012], za pomocą odpowiednio skonstruowanych wzorców ekstrakcji, wyszukiwano sekwencji takich jak „wydawał się niewystarczająco X ” czy też „czułem się prawie X ”. Wyrazy pojawiające się w tekście na miejscu zmiennej X , jeśli są przymiotnikami lub przysłówkami, stawały się kandydatami do umieszczenia w leksykonie.

W przypadku chęci wykorzystania otrzymanego leksykonu do analizy wydźwięku na poziomie aspektów, za każdym razem konieczne jest dodatkowo określenie aspektu, do którego dane sformułowanie się odnosi. W pracy [Lu i in. 2011], dokumenty dzielono z wykorzystaniem parsera na człony, będące zazwyczaj fragmentami zdań. Człony były analizowane w celu ustalenia czy pojawiają się w nich odniesienia do zdefiniowanych aspektów poprzez sprawdzenie czy w członie pojawił się wyraz z listy nazw danego aspektu. W przypadku, gdy wykryto dwa aspekty, przypisywano człon do obu z nich. Znajdujące się w członach przymiotniki i przysłówki traktowano jako kandydatów do leksykonu. Następnie, człony takie podlegały filtrowaniu na podstawie określonych kryteriów. Przykładowo, odrzucano człony ze zdań, w których pojawiały się słowa wskazujące na to, że ich znaczenie zależy od jakiegoś szerszego kontekstu, takie jak „jednakże”, gdyż w takim wypadku interpretacja mogłaby być błędna [Lu i in. 2011].

Jak wcześniej wspomniano, ekstrakcja sformułowań służących do wyrażania opinii w wielu przypadkach może odbywać się jednocześnie z identyfikacją polarności. Kolejne sekcje dotyczą już tego drugiego zagadnienia, jednak będą one również częściowo dotyczyły ekstrakcji sformułowań do leksykonu.

3.4.3 Identyfikacja polarności – metody słownikowe

Celem tego typu metod jest identyfikacja polarności sformułowań służących do wyrażania opinii (i często identyfikacja takich sformułowań) poprzez analizę istniejących zasobów leksykalnych, takich jak generyczne słowniki czy leksykalne bazy wiedzy [Liu 2011]. Zazwyczaj, metody te skutkują powstaniem generycznych, niedopasowanych do konkretnej domeny leksykonów, gdyż analizowane zasoby nie posiadają informacji umożliwiających ustalenie polarności w stosunku do różnych bytów i ich aspektów.

Analiza relacji semantycznych pomiędzy wyrazami

Liczne zasoby leksykalne posiadają informacje o relacjach semantycznych pomiędzy słowami. Przykładem takiego zasobu jest WordNet. Jest to dostępna w Internecie baza danych leksykalnych dla języka angielskiego⁷, przechowywana w postaci sieci semantycznej (hierarchicznej mapy pojęć) [Miller 1995]. Rzeczowniki, czasowniki, przymiotniki oraz przysłowki są tutaj pogrupowane w zestawy synonimów (tzw. synsety), z których każdy wyraża określone pojęcie. Zestawy takie są ze sobą powiązane za pomocą określonych relacji semantycznych i leksykalnych, poziomych oraz pionowych. Przykładami relacji występujących w WordNet mogą być „antonim” czy też „hiponim / hiperonim”.

Informacja o części mowy i relacji z innymi synsetami może być wykorzystywana w identyfikacji polarności sformułowań służących do wyrażania opinii oraz poszerzenia leksykonu o nowe sformułowania. Wyrazy, które znajdują się w jednym synsecie można traktować jako synonimy, natomiast pomiędzy synsetami, które odpowiadają przymiotnikom mogą istnieć m.in. relacje antonimii (przykładowo między „szybki” i „wolny”) oraz relacja „podobny do” (np. między „szybki” i „rozpędzony”). Relacje te można traktować jako wskazówki czy wyrazy w połączonych nimi synsetach mają polarność taką samą (w przypadku wyrazów z jednego synsetu lub relacji „podobny do”), czy też przeciwną (w przypadku antonimii). Posiadając początkową listę przykładowych słów o znanej polarności i listę przymiotników, dla których usiłowano przewidzieć orientację polarności, możliwe było przeanalizowanie, czy któryś z analizowanych przymiotników nie pozostaje w relacji antonimii bądź synonimii do któregoś z wyrazów z listy początkowej (listy sformułowań o znanej polarności). Przymiotniki, dla których w ten sposób określono orientację, poszerzały następnie zbiór początkowy [Hu i Liu 2004a].

Omawianą procedurę można wykorzystać zarówno do przypisywania polarności do znanych

⁷Istnieją również warianty dla innych języków, np. SłowoSieć dla języka polskiego

sformułowań z leksykonu, jak również do identyfikacji nowych sformułowań. Opisaną metodę można traktować jako klasyfikację. W pracy [Kim i Hovy 2004], przetestowano m.in. następującą formułę do przeprowadzenia opisanego wnioskowania:

$$\begin{aligned} \arg \max_c P(c|w) &= \arg \max_c P(c)P(w|c) = \\ &= \arg \max_c P(c) \frac{\sum_{i=1}^n \text{count}(\text{syn}_i, c)}{\text{count}(c)} \end{aligned} \quad (3.4)$$

gdzie:

- c jest kategorią polarności (jedną z dwóch: pozytywną lub negatywną),
- w jest wyrazem, dla którego staramy się ustalić orientację polarności,
- syn_i są to synonimy w .

Dla każdej z klas ustalone jest więc prawdopodobieństwo przynależności słowa do danej klasy na podstawie liczby synonimów wyrazu w należących do tej klasy. Do leksykonu były dodawane tylko te wyrazy, dla których prawdopodobieństwo przekraczało pewien ustalony poziom [Kim i Hovy 2004]. Opisywane podejście może zostać poszerzone o analizę wyrazów pozostających w większej odległości niż tylko jedna relacja [Godbole, Srinivasaiah i Skiena 2007]. Na przykład, w pracy [Kamps i in. 2004] analizowano odległość z wykorzystaniem relacji synonimii, gdzie orientację wyrazu określano na podstawie jego odległości od słów „dobry” (*good*) oraz „zły” (*bad*) według następującej formuły:

$$\text{EVA}(w) = \frac{d(w, \text{bad}) - d(w, \text{good})}{d(\text{good}, \text{bad})} \quad (3.5)$$

gdzie poprzez $d(a, b)$ określa się liczbę relacji synonimicznych, które dzielą wyrazy a oraz b . Otrzymana miara znajduje się w przedziale $[-1; 1]$ (-1 oznacza orientację negatywną, zaś 1 pozytywną) [Kamps i in. 2004].

Omawianą analizę można poszerzyć również o inne części mowy, np. czasowniki (wówczas jedyną analizowaną relacją może być synonimia) [Kim i Hovy 2004] czy rzeczowniki, gdzie dodatkowo można analizować relację hiperonimii, traktowaną jako utrzymującą tę samą orientację. Jednak, jak stwierdzono w pracy [Esuli i Sebastiani 2005], wykorzystanie relacji hiperonimii powoduje otrzymanie gorszych rezultatów.

Analiza definicji słów

Wiele zasobów leksykalnych, oprócz relacji pomiędzy słowami, zawiera także definicje słów. Dla każdego słowa może istnieć wiele definicji w zależności od tego, ile znaczeń dany wyraz posiada. Klasyfikacja takich definicji jest kolejnym podejściem do określania polarności słów. Wychodzi się tutaj z założenia, że słowa o podobnej polarności mają w pewien sposób podobne definicje i że z kolei same zawierają sformułowania o określonej orientacji semantycznej (przykładowo, definicja słowa o polarności pozytywnej będzie zawierała również sformułowania o takiej orientacji) [Esuli i Sebastiani 2005].

Określanie polarności w ten sposób można przeprowadzić za pomocą uczenia maszynowego. Przykładami uczącymi mogą być tu definicje sformułowań o znanej polarności. Na podstawie analizy definicji tworzona może być ich reprezentacja, na przykład w postaci wektora, gdzie kolejne elementy odpowiadają wyrazom ze słownika i ich wartości są ustalane zgodnie z miarą *TF-IDF*. W pracy [Esuli i Sebastiani 2005], przeprowadzono uczenie maszynowe dla tak skonstruowanych reprezentacji definicji i klasyfikację z wykorzystaniem naiwnego klasyfikatora Bayesa.

3.4.4 Identyfikacja polarności – metody wykorzystujące analizę korpusu

Metody wykorzystujące analizę korpusu mają na celu analizę dużej liczby recenzji napisanych w języku naturalnym, w celu wykrycia w nich sformułowań wyrażających opinie lub też w celu identyfikacji, jaką opinię te sformułowania wyrażają. Metody takie wymagają zazwyczaj dodatkowych informacji, takich jak:

- wcześniej przygotowanej listy przykładowych sformułowań o przypisanej polarności; w pracy [Turney i Littman 2003] noszą one nazwę sformułowań-paradygmatów, tak też dalej będą nazywane w tej pracy,
- przedstawionych w łatwo przetwarzalnej formie opinii o recenzowanych w korpusie przedmiotach (np. za pomocą punktacji w skali od 1 do 5 - patrz sekcja 2.4),
- adnotacje w korpusie, za pomocą których oznaczono sformułowania wyrażające opinie, w celu przeprowadzenia uczenia nadzorowanego.

Spójność wydźwięku

Klasycznym podejściem z zakresu metod wykorzystujących analizę korpusu jest tzw. spójność wydźwięku. Została ona zaprezentowana po raz pierwszy w pracy [Hatzivassiloglou i McKeown

1997]. Podejście to wykorzystuje analizę spójników pomiędzy przymiotnikami napotkanymi w tekście (konieczne jest tu posiadanie korpusu z adnotacjami przypisującymi części mowy do wyrazów). Wykorzystuje się tu spostrzeżenie, że wyrazy połączone określonymi spójnikami mogą mieć bądź taką samą bądź przeciwną orientację. Przykładowo, jeśli wyraz „błyskotliwy” ma pozytywne znaczenie oraz napotkamy sformułowanie „błyskotliwy i nieprzewidywalny” to z faktu, że wyrazy te łączy spójnik „i” możemy wywnioskować, że mają one taką samą, pozytywną orientację. Natomiast spójniki takie jak „choć”, „ale” czy „lecz” zmieniają polarność pomiędzy łączonymi sformułowaniami (np. „błyskotliwy, ale krótki”) [Hatzivassiloglou i McKeown 1997].

Na podstawie analizy wielu sformułowań połączonych spójnikami, pomiędzy każdą parą wyrazów można ustalić prognozowane podobieństwo ich orientacji. Grupy słów o takiej samej orientacji można następnie uzyskać poprzez analizę skupień takich wyrazów. W pracy [Hatzivassiloglou i McKeown 1997] przypisanie orientacji do skupień przeprowadzono opierając się o prostą regułę, że skupienie, w skład którego wchodzi wyrazy częściej występujące w tekście, odpowiada pozytywnej polarności. Innym sposobem określenia orientacji w tym podejściu jest wykorzystanie listy sformułowań służących do wyrażania opinii o znanej polarności. Na tej podstawie możliwe jest określenie, czy wyrazy, które są z tymi sformułowaniami połączone określonymi spójnikami, mają pozytywną lub negatywną orientację [Pang i Lee 2008].

Omawiane podejście może być rozszerzone na całe bloki wypowiedzi, w których wszystkie występujące przymiotniki mają taką samą orientację. Wnioskowanie takie wywodzi się z obserwacji, że często określona polarność opinii wyrażana jest w kilku następujących po sobie zdaniach. Zmiana orientacji rozpoznawana może być poprzez wystąpienie takich słów, jak „jednakże”, „pomimo to” czy też „niestety” [Kanayama i Nasukawa 2006].

Analiza współwystępowania sformułowań

Określenie częstości współwystępowania sformułowań jest kolejnym podejściem do identyfikacji polarności. Często stosowaną miarą jest tutaj tzw. informacja wspólna w punkcie PMI-IR (*Pointwise Mutual Information – Information Retrieval*). Jest ona określana na drodze odpytywania silników wyszukiwawczych i analizy liczby zwracanych przez nie wyników. Stosuje się tutaj następujący wzór:

$$\text{PMI}(\text{word}_1, \text{word}_2) = \log_2 \left(\frac{p(\text{word}_1 \& \text{word}_2)}{p(\text{word}_1)p(\text{word}_2)} \right) \quad (3.6)$$

Jako $p(\text{word}_1 \& \text{word}_2)$ rozumiemy tu liczbę wyników zwróconych w odpowiedzi na zapytanie

zawierające dwa sformułowania - word_1 oraz word_2 . W pracy [Turney i Littman 2003], sformułowania te łączono operatorem NEAR. Jedno ze sformułowań pochodzi ze zbioru paradygmatów, tzn. jego polarność jest znana. Ostateczną polarność sformułowania ustalano z wykorzystaniem następującego wzoru:

$$\begin{aligned} \text{SO-PMI}(\text{word}) = & \sum_{\text{pword} \in \text{Pwords}} \text{PMI}(\text{word}, \text{pword}) - \\ & - \sum_{\text{nword} \in \text{Nwords}} \text{PMI}(\text{word}, \text{nword}) \end{aligned} \quad (3.7)$$

Jako Pwords oznaczono kolekcję sformułowań-paradygmatów o pozytywnej, zaś jako Nwords o negatywnej polarności. Zaklasyfikowanie do odpowiedniej klasy słów zależało od tego czy ostateczny wynik był dodatni, czy ujemny.

Nadzorowanie odległe

Jak wspomniano w sekcji 1.2, nadzorowanie odległe jest to wykorzystanie równoczesnego przetwarzania dwóch kolekcji danych⁸, gdzie jedna z nich (strukturyzowana) traktowana jest jako kolekcja opisująca drugą (niestrukturyzowaną) [Broś 2013]. W przypadku identyfikacji polarności sformułowań, kolekcją strukturyzowaną mogą być opinie wyrażone na któryś ze strukturyzowanych formatów opisanych w sekcji 2.4, np. oceny punktowej lub w postaci list zalet i wad z predefiniowanymi opcjami. Kolekcją niestrukturyzowaną są tu opinie wyrażone w postaci wypowiedzi tekstowej, w których mogą występować sformułowania służące do wyrażania opinii z posiadanego leksykonu. Jako intuicyjne wyjaśnienie sposobu działania nadzorowania odległego w zadaniu identyfikacji polarności można przytoczyć obserwację z pracy [Liu 2011], w której stwierdzono, że jest wysoce nieprawdopodobne, żeby recenzenci używali w wypowiedzi tekstowej samych słów o negatywnej polarności w sytuacji, gdy za pomocą skali punktowej przyznali recenzowanemu bytowi najwyższą ocenę.

Prosty przykład nadzorowania odległego w zadaniu identyfikacji polarności zaprezentowano w pracy [Higashinaka, Prasad i Walker 2006]. W pracy tej wykorzystywano recenzje restauracji, w których, oprócz recenzji tekstowych, recenzenci wyrażali opinie za pomocą skali punktowej na temat pięciu aspektów. Za pomocą zdefiniowanych wzorców i leksykonu nazw aspektów identyfikowano sformułowania, za pomocą których recenzenci potencjalnie wyrażali opinie o tych aspektach. Następnie analizowano, jakie oceny w skali punktowej były nadawane aspektom, których nazwy

⁸Co inaczej można określić, tak jak w sekcji 1.2, jako jedną kolekcję, składającą się z dwóch części.

pojawiały się w wyekstrahowanych frazach. Należy tu zwrócić uwagę, że informacja płynąca ze strukturyzowanej kolekcji często jest niejednoznaczna. Przykładowo, w omawianej pracy do frazy „bardzo dobry” (*very good*) w odniesieniu do aspektu „jedzenie” opisana metoda przypisywała wszystkie możliwe oceny od 2 do 5. Finalną ocenę dla frazy można nadać poprzez uśrednienie ocen nadanych dla wszystkich recenzji, w których dana fraza wystąpiła. W pracy [Rill i in. 2012] przypisywano w ten sposób polarność do sformułowań często pojawiających się w tytułach recenzji. Natomiast w artykule [Sangiorgi, Augello i Pilato 2014] opisano wykorzystanie nadzorowania odległego (ponownie na podstawie ocen punktowych towarzyszących wypowiedziom tekstowym) do analizy zmieniającej się polarności słów z zadanego leksykonu w zależności od kontekstu, w którym dane słowo wystąpiło. Jako kontekst potraktowano okno składające się z określonej liczby słów przed i po analizowanym słowie.

Identyfikacja polarności za pomocą nadzorowania odległego może również przebiegać z wykorzystaniem list zalet i wad. W pracy [Kaji i Kitsuregawa 2007] zaproponowano metodę, w której przetwarzano listy bez predefiniowanych opcji. Opinie były tu wyrażane w postaci bardzo krótkich tekstów, dodanych albo do listy zalet, albo do listy wad. W związku z tym, do każdego z takich tekstów była przypisana jedna z dwóch polarności: pozytywna (jeśli tekst znajdował się w liście zalet) lub negatywna (jeśli znajdował się w liście wad). Posiadając wiele takich list zalet i wad, możliwe jest przeprowadzenie ich analizy i np. określenia, ile razy poszczególne sformułowania z leksykonu pojawiały się w listach zalet, a ile razy w listach wad. W pracy [Kaji i Kitsuregawa 2007], dla każdego sformułowania za pomocą testu χ^2 określano czy dane pozyskane w trakcie analizy pozwalają na przypisanie polarności do danego sformułowania. Podobne rozwiązanie zaprezentowano w pracy [Broś 2013], co dokładniej zostanie opisane w sekcji 7.1.3 na stronie 135.

Bardzo ciekawe podejście, które również można zaliczyć do nurtu nadzorowania odległego, zaprezentowano w pracy [Li i in. 2014]. Jako kolekcję opisującą recenzje określonych dóbr wyrażone w postaci wypowiedzi tekstowych potraktowano dane dotyczące ceny i poziomu sprzedaży tych dóbr. Kierowano się tu intuicją, że recenzje konsumenckie mogą wpłynąć na zainteresowanie konsumentów określonym dobrem, a tym samym znajomość danych sprzedażowych może pozwolić przewidzieć, jaki wydzźwięk mogą mieć recenzje tych dóbr. Przykładowo, jeśli dobro jest częściej kupowane, to prawdopodobnie jest ono lepiej oceniane. Metoda przypisywała więc polarność do słów występujących w recenzjach produktów na podstawie poziomu sprzedaży tych produktów. Wnioskowanie przeprowadzone jest tu z wykorzystaniem estymacji modeli ekonometrycznych.

Podejścia hybrydowe

Omówione powyżej podejścia zazwyczaj skupiają się na wykorzystaniu pojedynczego źródła informacji, na podstawie którego przeprowadza się ekstrakcję sformułowań służących do wyrażania opinii lub identyfikację ich polarności. Źródła takie mogą jednak oczywiście być łączone. W pracy [Lu i in. 2011] połączono następujące informacje, zwane tutaj sygnałami:

- generyczny (niezależny od domeny) leksykon sformułowań służących do wyrażania opinii,
- ogólną ocenę przypisaną do recenzji (w skali punktowej od 1 do 5),
- tezaurus, zawierający relacje synonimiczne i antonimiczne pomiędzy wyrazami,
- heurystyki lingwistyczne, takie jak te opisane w sekcji 3.4.4.

Połączenie tak różnorodnych sygnałów może być problematyczne. W pracy [Lu i in. 2011] sprowadzono ten problem do zadania optymalizacji w programowaniu liniowym, gdzie minimalizowano różnicę pomiędzy znaną polarnością słowa, a jego polarnością założoną na podstawie otrzymanych sygnałów. Minimalizowana funkcja jest ważoną sumą funkcji dla poszczególnych sygnałów [Lu i in. 2011].

3.5 Drażenie opinii w tekstach w języku polskim

W badaniach prezentowanych w dalszych rozdziałach tej pracy, drażenie opinii będzie przeprowadzane dla tekstów w języku polskim. W związku z tym warto przeanalizować, jaki jest stan badań nad drażeniem opinii dla języka polskiego. Niniejsza sekcja zostanie poświęcona właśnie temu zagadnieniu.

3.5.1 Różnice pomiędzy językami a metody i zasoby analizy wydźwięku

Jak dla każdego zagadnienia związanego z przetwarzaniem języka naturalnego, język polski stanowi znacznie większe wyzwanie niż choćby język angielski, między innymi z powodu bogatej fleksji, większej swobody w szyku wyrazów w zdaniu [Przepiórkowski 2007] oraz znacznie mniejszej liczby dostępnych zasobów lingwistycznych. Z tych powodów, przeniesienie metod drażenia opinii wypracowanych dla języka angielskiego na grunt języka polskiego jest trudnym zadaniem [Kowalska, Cai i Wade 2012]. Z drugiej strony, bogata fleksja języka polskiego umożliwia uży-

skanie dodatkowej informacji. Przykładowo, w pracy [Jędrzejewski i Morzy 2011] stwierdzono, że różne odmiany konkretnych wyrazów częściej pojawiają się w tekstach o określonym wydźwięku.

Metody drażenia opinii różnią się od siebie możliwością przeniesienia na grunt innych języków niż ten, dla którego zostały przygotowane. Przykładowo, podejścia statystyczne (np. uczenie maszynowe) mogą działać dla dowolnych języków naturalnych. W przypadku wykorzystania tego podejścia dla nowego języka konieczne jest jedynie pozyskanie odpowiednio zaadnotowanego korpusu i przeprowadzenie nauki modelu statystycznego. Natomiast w przypadku przeprowadzania na tekstach głębokiej analizy składniowej, możliwość przeniesienia rozwiązania na grunt innego języka staje się znacznie trudniejszym zadaniem.

Innym podejściem jest wykorzystanie automatycznego tłumaczenia tekstu do języka, dla którego istnieją metody drażenia opinii bądź zasoby ułatwiające to zadanie. Podejście takie zaprezentowano w pracy [Denecke 2008], gdzie tłumaczono dokumenty na język angielski, które następnie analizowano z wykorzystaniem SentiWordNetu. Innym wariantem takiego podejścia jest przetłumaczenie dostępnych zasobów językowych (np. zaadnotowanych korpusów) na język, dla którego ma być przeprowadzana analiza wydźwięku [Wan 2009]. W pracy [Balahur i Turchi 2012] wykazano, że spadek poprawności analizy wydźwięku, nie jest na tyle duży, aby wykluczyć możliwość zastosowania opisanego podejścia w praktyce. W pracy [Hogenboom i in. 2014] zaproponowano usprawnienie tego podejścia, poprzez wykorzystanie zasobów leksykalnych w analizowanych językach oraz relacji pomiędzy znaczeniami wyrazów, które w wielu przypadkach są definiowane pomiędzy takimi zasobami. W pracy generowano słownik sformułowań służących do wyrażania opinii dla języka holenderskiego na podstawie relacji pomiędzy SentiWordNetem, czyli leksykonem sformułowań służących do wyrażania opinii dla języka angielskiego, a angielskim WordNetem, a następnie relacji pomiędzy angielskim WordNetem, a jego odpowiednikiem dla języka holenderskiego.

3.5.2 Opublikowane prace z zakresu analizy wydźwięku w polskich tekstach

Badania nad analizą wydźwięku dla języka polskiego prowadzone są między innymi w Instytucie Podstaw Informatyki PAN. Prawdopodobnie pierwszymi opublikowanymi pracami z tego zakresu przeprowadzonymi w tym ośrodku były [Buczynski i Wawer 2008a] i [Buczynski i Wawer 2008b]. Przedstawiono w nich analizę wydźwięku całych dokumentów tekstowych (bez podziału na opisywane byty i aspekty). Wykorzystano w tym celu generyczny leksykon sformułowań wyrażających opinie przygotowany przez firmę Zetema. W badaniach porównano skuteczność metody typu *bag*

of words oraz metody wykorzystującej płytkie parsowanie języka z wykorzystaniem programu Spejd⁹. Autorzy przygotowali w sposób półautomatyczny zestaw reguł reprezentujących różne napotymane struktury wyrażające emocje, jak na przykład negację czy wyzerowanie (na przykład sformułowanie *zero defektów*), poprzez najpierw ekstrakcję często występujących bigramów (czyli 2-gramów), a następnie ręczne skonstruowanie reguł na podstawie ich analizy. Autorzy osiągnęli skuteczność metody na poziomie od 70% do 82%.

Dalsze prace w zakresie analizy wydźwięku prowadził wszystkim Aleksander Wawer. W artykule [Wawer 2011] zaprezentowano eksperyment, w którym wcześniej wybrane fragmenty tekstu podlegały klasyfikacji na podstawie ich podobieństwa w celu wykrycia fragmentów oceniających określone aspekty recenzowanych przedmiotów. Dla potrzeb nauki algorytmu, wykorzystano ręcznie zaadnotowany korpus, zaś sama klasyfikacja odbywała się z wykorzystaniem maszyny wektorów wspierających. Poprawność klasyfikacji wahała się między 30% a 53%. Z kolei w pracy [Wawer 2012] autor zaprezentował wynik eksperymentów nad doбором słów-paradygmatów dla opisywanej w sekcji 3.4.4 metody SO-PMI. Autor zwrócił uwagę na fakt, że często słowa takie dobierane są w sposób losowy. W celu doboru paradygmatów przetestowano rozkład SVD oraz analizę skupień, operujące na symetrycznej macierzy współwystąpień wyrazów, będących kandydatami na słowa wyrażające opinie. Wyrazy do tej macierzy wybrano z korpusu na drodze ekstrakcji słów (odpowiednich części mowy) występujących po sformułowaniach pochodzących z ręcznie przygotowanej listy wzorców ekstrakcji, takich, jak *wydawał się nie dość X, czułem się zbyt Y, był nadmiernie Z* itp.

W artykule [Kowalska, Cai i Wade 2012] zaprezentowano badania z zakresu drażenia opinii w krótkich wypowiedziach (od 10 do 100 słów) na forach pod artykułami na portalu tvn24.pl (określano wydźwięk całej wypowiedzi, bez podziału na aspekty itp.). Wykorzystano w tym celu uczenie maszynowe, a dokładniej naiwny klasyfikator Bayesa. Model oparto o wagę pojedynczych termów w poszczególnych dokumentach. W eksperymentach przetestowano cztery różne sposoby określania wagi termów. Przed nauką modelu przetworzono również teksty, m.in. poprzez określanie podstawy słowotwórczej wyrazów.

Kolejnym artykułem jest [Jędrzejewski i Morzy 2011]. Zaprezentowano tutaj podejście podobne do opisanego w [Dave, Lawrence i Pennock 2003], to jest wykorzystujące listę słów o polarności pozytywnej i negatywnej, określające wydźwięk całego dokumentu na podstawie tego, których słów jest więcej. Zmodyfikowano to podejście w ten sposób, że poszczególne wyrazy

⁹<http://zil.ipipan.waw.pl/Spejd/>

miały polarność o różnym stopniu intensywności w zależności od tego jak często pojawiały się w dokumentach o określonym wydźwięku. Eksperymenty przeprowadzono na wypowiedziach z trzech portali umożliwiających nadawanie punktacji w recenzjach, dzięki czemu możliwe było automatyczne określenie wydźwięku całego dokumentu, co było konieczne do nauki wydźwięku poszczególnych słów. Wprowadzona przez autorów modyfikacja wykazała poprawę w stosunku do podstawowego wariantu zaprezentowanego w [Dave, Lawrence i Pennock 2003].

W pracy [Lula i Wójcik 2011] opisano podejście do drażenia opinii w języku polskim z wykorzystaniem metod tłumaczenia maszynowego z języka polskiego do sztucznie stworzonego języka ODAL (*Opinion Description Artificial Language* - Sztuczny Język do Opisu Opinii). Język ten był w stanie reprezentować m.in. opinię dotyczącą określonego aspektu, informację o tym czy przedmiot jest tani czy drogi, informację o wyposażeniu przedmiotu w pewien aspekt lub brak takiego wyposażenia i ogólny wydźwięk. Następnie ręcznie przygotowano odwzorowania fragmentów wypowiedzi w języku polskim na odpowiednie sformułowania w ODAL. Na tej podstawie przeprowadzono standardową naukę modelu do automatycznego tłumaczenia tekstów. Niestety, w pracy [Lula i Wójcik 2011] nie poddano wypracowanej metody żadnej formalnej ocenie.

3.5.3 Rozwiązania komercyjne dotyczące analizy wydźwięku dla języka polskiego

Warto w tym miejscu przeanalizować również ofertę komercyjnych podmiotów zajmujących się zagadnieniami związanymi z drażeniem opinii. Firmy te świadczą przede wszystkim usługi monitorowania reputacji marki w Internecie (z uwzględnieniem monitorowania serwisów społecznościowych). Ciekawą publikacją ukazującą stan zaawansowania systemów proponowanych przez te firmy na rok 2011 jest raport „Emotrack: Marki RTV w wypowiedziach internautów”, opublikowany przez firmy Think Kong i Zetema¹⁰. Poddano w nim analizie wypowiedzi internautów pochodzące z prawie 6000 serwisów internetowych marek i produktów z branży RTV. Badania miały na celu określenie:

- opinii o markach RTV oferujących szeroką gamę produktów (dla ośmiu wybranych marek),
- opinii o produktach,
- uniwersalnych aspektów ocenianych w domenie urządzeń RTV,

¹⁰<http://www.emotrack.pl>

- opinii o rozpoznanych aspektach.

W trakcie badań wykorzystywano odpowiednio skonstruowane słowniki, wzorce ekstrakcji oraz analizę składniową. Niestety, metody te nie zostały dokładnie opisane, nie opublikowano również oceny poprawności ich działania. Dostępne są natomiast wyniki analizy wydźwięku, dzięki którym można poznać to, w jaki sposób w polskim Internecie oceniane są poszczególne marki, produkty i ich aspekty.

3.6 Zidentyfikowane braki w istniejących metodach

Z zaprezentowanej analizy literatury wynika, że zagadnieniu drażenia opinii poświęcono już wiele badań. Jednocześnie, wskazuje się na liczne wyzwania stojące przed dziedziną. Wynikają one m.in. z bogactwa języka, w którym pisane są recenzje. Spośród tych trudności warto wymienić:

- przetwarzanie pytań oraz zdań warunkowych – tego typu zdania w wielu przypadkach nie zawierają opinii nawet wówczas, gdy znajdują się w nich słowa zazwyczaj wyrażające opinie. Przykładowo, pytanie „Czy ktoś zna dobrą kamerę firmy XYZ” nie wyraża żadnej opinii, pomimo tego, że występuje w nim słowo „dobrą”. Z drugiej strony, w pytaniu „Czy ktoś mógłby mi powiedzieć, dlaczego ten aparat wciąż się psuje?” zawiera negatywną opinię o produkcie. Właściwe rozróżnienie takich przypadków bywa bardzo kłopotliwe [Liu 2011],
- wielowyzrazowe sformułowania wyrażające opinię nie wprost, gdzie nie pojawiają się ani słowa wyrażające opinię, ani odnoszące się wprost do aspektu [Hu i Liu 2004a]; przykładowo, sformułowanie *Telefon nie mieści się łatwo w kieszeni* odnosi się do rozmiaru (lub też kształtu) telefonu w negatywny sposób, jednak automatycznie wywnioskowanie tego faktu ciągle pozostaje znaczącym wyzwaniem [Moghaddam i Ester 2012],
- recenzje porównujące – w wielu recenzjach mogą pojawić się odniesienia do konkurencyjnych produktów. W tym przypadku odróżnienie tego, które sformułowania odnoszą się do którego produktu wymaga bardzo głębokiego zrozumienia wypowiedzi. Przykładem trudności związanych z tym zagadnieniem jest interpretacja zaimków (tzw. problem ujednoznaczniania koreferencji, ang. *coreference resolution*), [Liu 2011] (przykładowo ”Telefon A jest lepszy od B, natomiast nieco gorzej od niego wygląda”),
- sarkazm – w recenzjach nieraz można natknąć się na sarkazm czy ironię, których występowanie zazwyczaj odwraca znaczenie poszczególnych słów występujących w tekście (przykładowo

w zdaniu „Co za świetny aparat, zepsuł się tydzień po zakupie”) [Chaovalit i Zhou 2005; Liu 2011],

- uczenie maszynowe z wykorzystaniem nadzorowania odległego – w wielu przypadkach, założenia dotyczące posiadania dwóch kolekcji, w których wyrażone zostały te same opinie, nie jest spełnione. Dokładniej ten problem zostanie opisany w sekcji 4.1.

3.7 Podsumowanie

W rozdziale zaprezentowano wyniki analizy literatury z zakresu wybranych zagadnień związanych z drażeniem opinii. W rozdziale poświęcono uwagę metodom analizy wydźwięku, określaniu aspektów ocenianych w recenzjach oraz metodom konstrukcji leksykonów sformułowań służących do wyrażania opinii. Następnie w sekcji 3.5 przeanalizowano stan rozwoju drażenia opinii w tekstach w języku polskim, na końcu zaś przedstawiono zidentyfikowane trudności, jakie wciąż stoją przed dziedziną analizy wydźwięku.

Z przedstawionej analizy wynika, że zagadnieniu drażenia opinii w tekstach angielskojęzycznych poświęcono już wiele badań. Jednocześnie, badania dla języka polskiego są na znacznie wcześniejszym stadium zaawansowania. Niezależnie jednak od języka, istnieje wciąż wiele nierozwiązanych problemów, które utrudniają przeprowadzanie drażenia opinii.

Rozdział 4

Wykorzystanie podsumowań opinii o produktach i usługach do identyfikacji polarności – zarys podejścia

Celem rozdziału jest zaprezentowanie koncepcji autorskiej metody nauki polarności sformułowań wykorzystywanych przez recenzentów do oceny różnych aspektów produktów i usług. Koncepcja ta łączy obserwacje poczynione w trakcie analizy literatury i analizy dziedziny recenzji konsumenckich, wyniki których zostały zawarte w poprzednich dwóch rozdziałach pracy. Głównymi źródłami inspiracji przy formułowaniu koncepcji była analiza różnych sposobów wyrażania opinii oraz identyfikacja problemów związanych z automatycznym przypisywaniem polarności do sformułowań wyrażających opinie.

W rozdziale wyjaśnione zostanie obrane podejście, zaprezentowane zostaną wybrane warianty metody (zależne od danych, na których metoda ma działać oraz scenariuszy wykorzystania), a następnie sformułowane zostaną przyjęte założenia dla metody. Należy zaznaczyć, że w tym rozdziale skupiono się wyłącznie na metodzie bezpośrednio powiązanej z tezą pracy i pominięto szerszy kontekst innych etapów przetwarzania, które konieczne są do umożliwienia przeprowadzenia identyfikacji polarności analizy wydźwiku. Szerszy opis metody, uzupełniający obszary pominięte w tym rozdziale, zostanie przedstawiony w rozdziale 6.

Obszerne fragmenty niniejszego rozdziału opublikowano w artykułach [Bernatowicz i Małyszko 2014; Małyszko 2013; Małyszko i Filipowska 2013].

4.1 Nadzorowanie odległe a identyfikacja polarności – problem z założeniami

Wiele z podejść opisanych w rozdziale 3 można zaliczyć do nurtu nadzorowania odległego, opisanego w sekcji 3.4.4. Podejście to wymaga spełnienia przez posiadany korpus pewnych specyficznych założeń, które warto przytoczyć i przeanalizować.

Jak wspomniano, w nadzorowaniu odległym konieczne jest dysponowanie dwiema kolekcjami: niestrukturyzowaną oraz strukturyzowaną, traktowaną jako opis tej pierwszej. Przykładami takich kolekcji są:

- w pracy [Rill i in. 2012] na kolekcję niestrukturyzowaną składały się tytuły recenzji nadawane przez ich autorów, natomiast kolekcją strukturyzowaną, służącą do opisu niestrukturyzowanej, były przypisane przez recenzentów oceny punktowe,
- w artykule [Higashinaka, Prasad i Walker 2006] wykorzystano korpus, w którym do każdej recenzji tekstowej przypisanych jest kilka ocen punktowych, odpowiadających zdefiniowanemu aspektowi produktu; kolekcja strukturyzowana składa się z ocen punktowych przypisanych do danej recenzji,
- w pracach [Broś 2013; Kaji i Kitsuregawa 2007] wykorzystywano kolekcję list zalet i wad, gdzie poszczególne zalety i wady miały postać dowolnych wypowiedzi tekstowych wyrażanych przez recenzentów; te wypowiedzi były traktowane jako kolekcja niestrukturyzowana, natomiast kolekcją strukturyzowaną były informacje o tym, czy dana wypowiedź znalazła się na liście zalet czy na liście wad.

W każdym z opisanych podejść konieczne jest, aby opinia została wyrażona przez tego samego recenzenta i w tym samym momencie na więcej niż jeden sposób – wówczas możliwe jest równoczesne przetwarzanie obu kolekcji w celu przeprowadzenia, przykładowo, nauki polarności sformułowań użytych w kolekcji niestrukturyzowanej. Nie zawsze jednak dysponujemy takimi danymi. W wielu przypadkach, opinie wyrażane są tylko na jeden sposób, przykładowo tylko za pomocą ocen punktowych (np. oceny filmów nadawane przez użytkowników portalu `filmweb.pl`) lub tylko w postaci wypowiedzi tekstowej (np. opinie wygłaszane w wypowiedziach zamieszczanych na portalach społecznościowych). Dodatkowo, nawet wówczas, gdy dysponujemy takimi dwiema kolekcjami (strukturyzowaną i niestrukturyzowaną, gdzie w obu kolekcjach posiadamy

opinie wyrażone przez tych samych recenzentów), kolekcje te mogą być niekompletne. Przykład takiej sytuacji opisany zostanie w sekcji 4.3.2.

Celem proponowanej metody jest zniesienie wspomnianych powyżej ograniczeń stosowania nadzorowania odległego. Poniżej wyjaśniono sposób, w jaki proponowane rozwiązanie ma to osiągnąć.

4.2 Zarys proponowanej metody

Podejście proponowane w pracy polega na identyfikacji polarności sformułowań służących do wyrażania opinii na podstawie analizy tego, jakie produkty, usługi i aspekty są za pomocą tych sformułowań oceniane, wykorzystując przy tym informacje dotyczące sposobu postrzegania tych produktów i usług przez wielu recenzentów. Wnioskowanie przeprowadzane w podejściu proponowanym w pracy można zobrazować za pomocą następującego scenariusza ze świata rzeczywistego.

Założmy, że pewna osoba (nazwijmy ją Alice) pochodząca z Anglii jest w zagranicznej podróży służbowej w Polsce. Alice mieszkała wcześniej w Polsce przez dwa lata i zdołała przez ten czas w pewnym stopniu opanować język polski. Alice uczestniczy w spotkaniu, którego uczestnicy mówią po polsku. Poziom znajomości języka polskiego przez Alice pozwala jej zrozumieć większą część rozmowy.

Na spotkaniu Robert prowadzi prezentację. W jej trakcie Robert wyraził opinię o pewnej usłudze za pomocą słowa, którego Alice nie zna. Przykładowo, Robert powiedział zdanie: „Obsługa w liniach lotniczych XYZ jest marna”. Alice nigdy nie słyszała słowa „marna” i nie jest w stanie ustalić znaczenia tego wyrazu. Jednak tak się składa, że Alice słyszała kilka razy od innych osób bardzo negatywne opinie o jakości obsługi we wspomnianych liniach lotniczych. Innymi słowy, Alice posiada pewną wiedzę na temat sposobu postrzegania pewnego dobra przez innych. Na tej podstawie, Alice podejrzewa, że słowo „marna” służy do wyrażania negatywnych opinii (przynajmniej w kontekście obsługi). Odtąd, gdy Alice usłyszy ten wyraz, będzie już miała pewną intuicję dotyczącą tego, co ten wyraz może oznaczać.

Stykając się przez dłuższy czas z podobnymi sytuacjami (również np. podczas oglądania programów telewizyjnych w obcym języku) ludzie mogą poszerzać znajomość określonego języka. Oczywiście, nauka języka odbywa się również na wiele innych sposobów, jednak opisana metoda również może okazać się przydatna podczas tego procesu.

Proponowane rozwiązanie ma działać w analogiczny sposób. Liczba recenzji konsumenckich w

Internecie oraz fakt, że wiele z nich ma format umożliwiający łatwe ich przetwarzanie sprawiają, że w wielu przypadkach możliwe jest automatyczne pozyskanie informacji o sposobie postrzegania danego produktu lub danej usługi przez wielu recenzentów. Tak pozyskana informacja może być wykorzystana jako wskazówka w procesie nauki polarności słów wykorzystywanych w wypowiedziach tekstowych, w których znajdują się opinie o danym produkcie lub usłudze.

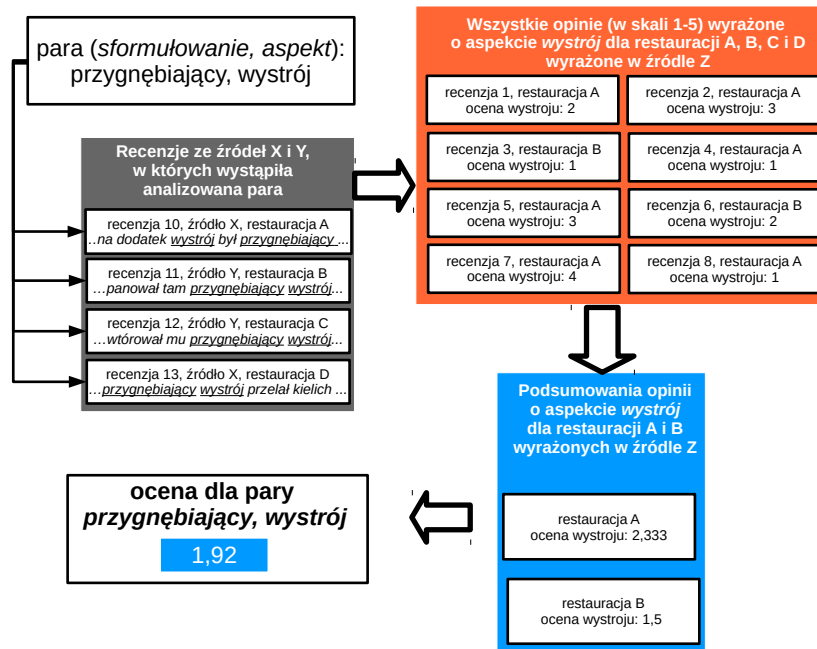
Proponowaną metodę można zaliczyć do nurtu nadzorowania odległego. Ponownie, mamy tu dwie kolekcje: kolekcję niestrukturyzowaną, złożoną z wypowiedzi tekstowych oraz strukturyzowaną, złożoną z opinii o produkcie lub usłudze wyrażonych w przetwarzalny sposób. Fundamentalną różnicą w proponowanym rozwiązaniu w stosunku do opisanych wcześniej podejść jest to, że obie kolekcje (strukturyzowana i niestrukturyzowana) mogą pochodzić od różnych recenzentów, z różnego momentu w czasie, a nawet z różnych źródeł. Takie rozluźnienie założeń może pozwolić na przeprowadzenie wnioskowań w sytuacjach, w których istniejące podejścia z zakresu nadzorowania odległego nie mogą być stosowane. Teza, postawiona w prezentowanej pracy (przytoczona we Wprowadzeniu), dotyczy tak opisanego wnioskowania i eksperymenty zaprezentowane w dalszych rozdziałach pracy miały za zadanie jej wykazanie.

4.3 Warianty proponowanej metody

Podejście zarysowane w poprzedniej sekcji można uszczegółowić, tworząc osobne warianty metody, możliwe do wykorzystania w różnych scenariuszach oraz dla różnych danych, na których metoda ma operować. Tym samym należy zaznaczyć, że proponowane podejście może zostać wykorzystane w szerokim zakresie zastosowań. Poniżej zaproponowano trzy takie warianty.

4.3.1 Wariant I

Założmy, że mamy do czynienia z systemem, którego zadaniem jest analiza opinii o restauracjach z pewnego obszaru geograficznego, publikowanych w postaci wypowiedzi tekstowych w określonych źródłach internetowych X i Y (są to przykładowo portale społecznościowe) i informowanie właścicieli restauracji o wpisach, w których użytkownicy negatywnie odnoszą się do danej restauracji. Aby system poprawnie mógł spełniać swoją funkcję, konieczne jest posiadanie przez niego leksykonu sformułowań, za pomocą których recenzenci mogą wyrażać swoje opinie o różnych aspektach restauracji (np. obsłudze, jakości jedzenia, wystroju, atmosferze itp.; założmy, że system ma zdefiniowaną kolekcję aspektów, opinie o których mają być poddane analizie). Sformułowania te muszą



Rysunek 4.1: Proces wyliczania statystyki umożliwiającej przeprowadzenie identyfikacji polarności na podstawie podsumowań wielu opinii o recenzowanym bycie dla pierwszego wariantu proponowanej metody. Źródło: opracowanie własne

mieć dodatkowo przypisane polarności w zależności od tego, do którego aspektu recenzowanego dobra się odnoszą. W związku z tym, zanim możliwe będzie przeprowadzenie analizy wydźwięku, konieczne jest przypisanie polarności do par (*sformułowanie, aspekt*).

Założmy, że analizowany system posiada metodę, pozwalającą na wykrywanie występujących w tekście par (*sformułowanie, aspekt*), tak jak zaprezentowano to na rysunku 4.1, gdzie wykryto, że w czterech recenzjach występuje para (*przygnębiający, wystrój*). System nie posiada jednak informacji o tym, jaką polarność należy przypisać do tej pary.

Proponowana metoda ma pozwolić na identyfikację polarności dla analizowanej pary poprzez wykorzystanie dodatkowego portalu (nazwijmy go źródłem Z), w którym ocenie podlegają restauracje i ich aspekty, gdzie opinie są wyrażane w postaci punktowej. Dla każdej restauracji w źródle Z możliwe jest przygotowanie podsumowania wszystkich opinii o danej restauracji o określonych jej aspektach, możliwych do oceny na danym portalu. Podsumowanie opinii może mieć postać taką, jak zaprezentowano to na rysunku 4.2. Nazwijmy ten portal portalem referencyjnym. Założmy, że na portalu referencyjnym recenzenci mogą wyrażać opinie za pomocą skali punktowej o interesującym nas w omawianym przykładzie aspekcie, tj. o aspekcie *wystrój*. Zaproponowana metoda



Rysunek 4.2: Zrzut ekranu z portalu gastronauta.pl prezentujący przykład podsumowania opinii recenzentów o różnych aspektach wybranej restauracji. Dostęp 29 kwietnia 2014

identyfikuje polarność dla pary (*przygnębiający, wystrój*) zgodnie z następującą procedurą:

1. Dla każdej recenzji, w której występuje analizowana para, system określa jaka restauracja jest przedmiotem recenzji.
2. Ze źródła referencyjnego automatycznie pozyskiwane są recenzje zidentyfikowanych restauracji, czyli w tym przypadku A, B, C i D (załóżmy w tym momencie, że jesteśmy w stanie jednoznacznie rozpoznać, jakiej restauracji dotyczy recenzja, niezależnie od źródła, z którego ta recenzja pochodzi). W przykładzie pokazanym na rysunku 4.1, w źródle Z znajduje się sześć recenzji restauracji A i dwie recenzje restauracji B, natomiast restauracje C i D nie zostały w źródle Z ocenione.
3. System pobiera z portalu referencyjnego podsumowanie opinii wszystkich recenzentów o aspekcie „wystrój” dla analizowanych restauracji (w tym przypadku odnaleziono tylko recenzje restauracji A i B); na zaprezentowanym rysunku do obliczenia podsumowań wykorzystano średnią arytmetyczną.
4. Uśredniając otrzymane podsumowania dla wszystkich analizowanych restauracji, system generuje ostateczną wartość, odpowiadającą temu, jak oceniany jest w źródle referencyjnym aspekt *wystrój* dla tych restauracji, w recenzjach których w źródłach X i Y aspekt ten oceniony został z wykorzystaniem sformułowania *przygnębiający*.
5. Otrzymaną wartość można bezpośrednio przekształcić zgodnie z przyjętą regułą na polarność przypisaną do analizowanej pary lub też wykorzystać jako jedną z cech branych pod

uwagę podczas takiego przypisywania polarności.

Po wykonaniu opisanej procedury, system będzie w stanie przypisać polarność za każdym razem, gdy natknie się na recenzję, której wystąpiła para (*przygnębiający, wystrój*). Należy zauważyć, że po etapie identyfikacji polarności portal referencyjny nie jest już wykorzystywany. Warto zauważyć, że już dla danych wykorzystanych w powyższym przykładzie przeprowadzone wnioskowanie pozwoliło na osiągnięcie pewnych korzyści. O ile bowiem opinie o wystroju w restauracji A oraz restauracji B mogły zostać pobrane z portalu referencyjnego (choć były to opinie innych recenzentów niż tych, którzy wyrazili opinie w źródłach X i Y), to nie było możliwe automatyczne rozpoznanie opinii o tym aspekcie dla restauracji C oraz D. Po przeprowadzeniu opisanego wnioskowania, takie opinie, wyrażone w recenzjach 12 i 13, mogą zostać automatycznie rozpoznane.

4.3.2 Wariant II

Drugi wariant metody można zastosować do sytuacji, gdy dostępne są recenzje, w których jednocześnie wyrażano opinie w postaci recenzji tekstowych oraz list zalet i wad. W trakcie rozważań, jako przykład wykorzystany zostanie portal *cokupic.pl*. Na tym portalu, jednym z formatów wyrażania opinii jest lista zalet i wad z predefiniowanymi opcjami, specyficznymi dla danej kategorii produktów. Użytkownicy wyrażają swoją opinię o danym produkcie zaznaczając odpowiednie opcje. Jednocześnie, recenzenci wyrażają swoje opinie również w postaci wypowiedzi tekstowej.

Jak dla mnie telefonik okazał się całkiem fajny. Duży oraz bardzo wyraźny wyświetlacz to wielkie plusy tego modelu Samsunga. Aczkolwiek jedyna wada tego telefonu to to, że w łatwy sposób się zawiesza. Robi dość ciekawe i oryginalne zdjęcia, ponieważ ma bardzo dużo pikseli. Idealny do rozmów, słuchania muzyki oraz oglądania filmów. Pomimo jednej wady jaką posiada i tak polecam.

+ Plusy

- długa żywotność baterii

- dobry aparat fotograficzny

- Minusy

Rysunek 4.3: Przykład recenzji wyrażonej jednocześnie w postaci opisu tekstowego i listy zalet i wad. Część informacji powtarza się w obu częściach recenzji, jednak wiele opinii wyrażonych jest tylko w jednej z tych części. Źródło: <http://cokupic.pl/produkt/Samsung-Galaxy-S-II-GT-i9100> (dostęp 29 kwietnia 2013)

Prosty przykład nadzorowania odległego przy tak opisanym korpusie, niewykraczający poza dotychczasowe podejścia, zaprezentowano na rysunku 4.3. Jest to podejście podobne do zapre-

zentowanego w pracy [Higashinaka, Prasad i Walker 2006]. Wnioskowanie jest tu przeprowadzane w następujący sposób. Jeśli w tekście pojawi się nazwa pewnego aspektu wraz ze słowem o nieznannej polarności, a jednocześnie w liście zalet i wad oznaczono dany aspekt jako zaletę (bądź wadę) produktu, to do tego słowa należy przypisać polarność pozytywną (lub negatywną). Na rysunku 4.3, dzięki nadanej zalecie „dobry aparat fotograficzny” (założmy, że w systemie znajduje się informacja o tym, że zaleta ta dotyczy aspektu „aparat fotograficzny”) łatwo jesteśmy w stanie ustalić, że odnoszące się w tekście do słowa „zdjęcia” (o którym system również wie, że odnosi się do tego aspektu) wyrazy „ciekawe” i „oryginalne” mają prawdopodobnie w kontekście tego aspektu produktu pozytywną polarność. Na podstawie analizy wielu wystąpień omawianych słów dla aspektu „aparat fotograficzny” można więc określić polarność, jaka powinna zostać przypisana do tych słów dla omawianego aspektu.

Należy jednak zwrócić uwagę na fakt, że recenzenci, mając do dyspozycji jednocześnie możliwość wyrażenia recenzji w postaci tekstowej oraz za pomocą listy zalet i wad, często nie zamieszczają w nich tych samych informacji. W opisie tekstowym recenzent może skupić się na innych aspektach niż te, które ocenił na liście zalet i wad, przykładowo z tego powodu, że traktuje oba sposoby wyrażania opinii jako uzupełniające się i nie chce podawać tej samej informacji dwa razy. W związku z tym, w tej samej recenzji, w jej częściach wyrażonych za pomocą różnych sposobów, mogą znajdować się te same informacje, ale nie muszą. Przykładowo, na rysunku 4.3, w tekście nie pojawiła się żadna wzmianka o żywotności baterii telefonu (pomimo zaznaczenia pozycji „długa żywotność baterii” na liście zalet), natomiast na liście zalet i wad recenzent pominął ocenę rozmiaru i rozdzielczości wyświetlacza (pomimo faktu, iż opisał te aspekty w wypowiedzi tekstowej, a istnieje odpowiednia zaleta do wyboru z listy zalet w trakcie tworzenia recenzji). Z tego powodu, w przypadku analizy jedynie pojedynczych recenzji, tracone jest wiele informacji zawartych w opiniach wyrażonych za pomocą listy zalet i wad. Tym samym, przeprowadzone na analizowanym korpusie recenzji wnioski z zakresu nadzorowania odległego nie jest doskonałe.

Opisany problem można rozwiązać korzystając z wniosku zaproponowanego w sekcji 4.2. Zgodnie z zaproponowanym podejściem, system powinien wykorzystywać podsumowania opinii o produktach, wygenerowane na podstawie wielu recenzji dotyczących tego samego produktu, jako reprezentację jego postrzegania przez recenzentów. Przy założeniu, że możliwe zalety i wady pochodzą z predefiniowanej listy, możliwe jest łatwe przygotowanie takiego podsumowania w rozbiciu na poszczególne aspekty (takie podsumowania są generowane i prezentowane użytkownikom w istniejących systemach, patrz rysunek 2.3 na stronie 26, proponowano je również m.in. w pracy

Aspekt	Zalet	Wad
obsługa	1	0
liczba funkcji	0	16
rozmiar wyświetlacza	0	18
rozdzielczość wyświetlacza	0	5
żywołność baterii	3	11
jakość dźwięku	1	6
cena	12	0

Tabela 4.1: Podsumowanie przyznanych zalet i wad (na listach zalet i wad) dla wybranych aspektów telefonu komórkowego Alcatel-ot311 na podstawie strony <http://cokupic.pl/produkt/Alcatel-alcatel-ot311>

[Liu, Hu i Cheng 2005]). Otrzymane podsumowanie można zaprezentować w postaci tabeli, gdzie w poszczególnych wierszach przechowywane są informacje o tym, ilu recenzentów uznało dany aspekt jako zaletę, a ilu jako wadę produktu. Przykład zaprezentowano w tabeli 4.1. Wynika z niej m.in., że aspekt „żywołność baterii” w telefonie komórkowym, którego dotyczyły recenzje, został przez 3 recenzentów oceniony pozytywnie, a przez 11 negatywnie.

Sposób działania proponowanej metody zaprezentowano na rysunku 4.4 na stronie 82. Pokazano na nim schematycznie analizę przeprowadzoną dla przykładowej pary (*sformułowanie, aspekt*), tj. pary (*atrakcyjny, design*).

1. Zgodnie z diagramem przedstawionym na rysunku 4.4, para (*sformułowanie, aspekt*) w korpusie dokumentów była użyta w dwóch recenzjach: recenzji o identyfikatorze 1, w której oceniany był produkt A oraz recenzji 2 dotyczącej produktu B (załóżmy, że dysponujemy odpowiednią metodą, która pozwala na identyfikowanie takich par w analizowanych recenzjach). Recenzje te znajdują się na rysunku na ciemnoszarym polu. W każdej z tych recenzji, ich autorzy mogli wyrazić swoją opinię o różnych aspektach produktu (w tym również o aspekcie *design*) za pomocą listy zalet i wad. Na diagramie, przy każdej recenzji oznaczono, czy interesujący nas aspekt (*design*) został na takiej liście w danej recenzji oceniony. Wartość 1 przy etykietce zaleta lub wada oznacza, że aspekt był oceniony w dany sposób, natomiast 0 oznacza, że recenzent nie oznaczył danego aspektu na tej liście. Jeśli przy obu etykietkach znajduje się wartość 0 oznacza to, że recenzent w ogóle nie wyraził swojej opinii

o danym aspekcie na liście zalet i wad.

2. Dla każdej z recenzji, w których wystąpiła analizowana para, zidentyfikowano produkt, który był w niej oceniany (w naszym przypadku były to produkty A i B), a następnie pobierano wszystkie recenzje tego produktu z bazy.
3. Dla każdego ze zidentyfikowanych produktów, wszystkie opinie, które zostały wyrażone o aspekcie *design* podlegają podsumowaniu. Aspekt ten w recenzjach produktu A trzykrotnie został oceniony pozytywnie, a raz negatywnie (w dwóch przypadkach recenzenci nie wyrazili swojej opinii o tym aspekcie). Uzyskane w ten sposób podsumowanie podlega normalizacji do jedynki, tj. w podsumowaniu do zalet przypisujemy wartość $\frac{3}{3+1}$, natomiast do wad $\frac{1}{3+1}$. W związku z tym, dla produktu A otrzymaliśmy znormalizowane podsumowanie opinii o aspekcie *design*, mające postać *zaleta: 0,75, wada: 0,25*. Analogiczne, znormalizowane podsumowanie generowane jest dla produktu B.
4. Na podstawie przygotowanych podsumowań i recenzji, w których dane sformułowanie występowało, przygotowany jest wektor cech¹. Składa się on z dwóch głównych części. Pierwsza z nich (wyróżniona na szaro) obliczana jest na podstawie recenzji, w których dane sformułowanie wystąpiło. Dwie cechy, należące do tej części wektora, są obliczane analogicznie do istniejących metod nadzorowania odległego. Zliczane jest, ile razy w tych recenzjach dany aspekt był oznaczony jako zaleta, a ile razy jako wada, a następnie otrzymane wartości zalet i wad normalizowane są do jedności, zgodnie z następującymi wzorami:

$$zalety = \begin{cases} \frac{\sum_{r \in R} p_{r,a}}{\sum_{r \in R} p_{r,a} + \sum_{r \in R} n_{r,a}} & \text{jeśli } \sum_{r \in R} p_{r,a} + \sum_{r \in R} n_{r,a} > 0 \\ 0 & \text{jeśli } \sum_{r \in R} p_{r,a} + \sum_{r \in R} n_{r,a} = 0 \end{cases} \quad (4.1)$$

$$wady = \begin{cases} \frac{\sum_{r \in R} n_{r,a}}{\sum_{r \in R} p_{r,a} + \sum_{r \in R} n_{r,a}} & \text{jeśli } \sum_{r \in R} p_{r,a} + \sum_{r \in R} n_{r,a} > 0 \\ 0 & \text{jeśli } \sum_{r \in R} p_{r,a} + \sum_{r \in R} n_{r,a} = 0 \end{cases} \quad (4.2)$$

Zastosowano tu następujące oznaczenia:

- R to kolekcja wszystkich recenzji, w których pojawiła się dana para,
- $p_{r,a}$ ma wartość 1, jeśli w recenzji r aspekt a został oznaczony na liście zalet lub 0 jeśli nie został w niej oznaczony,

¹Posługujemy się tu terminem wektora cech w rozumieniu opisanym już w podsekcji 3.3.1

- $n_{r,a}$ ma wartość 1, jeśli w recenzji r aspekt a został oznaczony na liście wad lub 0 jeśli nie został w niej oznaczony.

Otrzymane wartości umieszczane są w wektorze i na diagramie opatrzone je etykietami *zalety* i *wady*.

Druga część wektora jest obliczana na podstawie znormalizowanych podsumowań opinii wszystkich recenzentów o aspekcie dla produktów A i B. Cechy obliczane w ten sposób nie były dotychczas wykorzystywane w identyfikacji polarności. Sumowane są wartości dla wszystkich produktów (w tym przypadku dwóch, A i B), które następnie ponownie podlegają normalizacji, zgodnie z następującymi formułami:

$$zalety_p = \begin{cases} \frac{\sum_{s \in S} p_{s,a}}{\sum_{s \in S} p_{s,a} + \sum_{s \in S} n_{s,a}} & \text{jeśli } \sum_{s \in S} p_{s,a} + \sum_{s \in S} n_{s,a} > 0 \\ 0 & \text{jeśli } \sum_{s \in S} p_{s,a} + \sum_{s \in S} n_{s,a} = 0 \end{cases} \quad (4.3)$$

$$wady_p = \begin{cases} \frac{\sum_{s \in S} n_{s,a}}{\sum_{s \in S} p_{s,a} + \sum_{s \in S} n_{s,a}} & \text{jeśli } \sum_{s \in S} p_{s,a} + \sum_{s \in S} n_{s,a} > 0 \\ 0 & \text{jeśli } \sum_{s \in S} p_{s,a} + \sum_{s \in S} n_{s,a} = 0 \end{cases} \quad (4.4)$$

W powyższych wzorach, oznaczenia mają następującą interpretację:

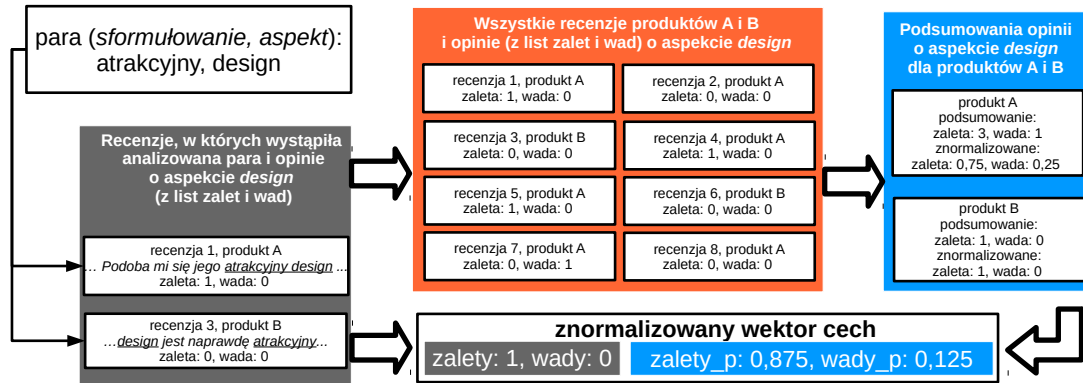
- S jest to kolekcja, na którą składają się podsumowania wszystkich opinii o poszczególnych produktach, których dotyczą recenzje należące do R ; w naszym przypadku znajdują się w tej kolekcji dwa podsumowania (dla produktu A i dla produktu B),
- $p_{s,a}$ to znormalizowana wartość, przypisana w podsumowaniu s zaletom,
- $n_{s,a}$ to znormalizowana wartość, przypisana w podsumowaniu s wadom.

Przykładowo, wartość *zalety_p* na rysunku 4.4 została obliczona w następujący sposób:

$$\frac{0,75 + 1}{0,75 + 1 + 0,25 + 0} = \frac{1,75}{2} = 0,875.$$

Analogiczne operacje przeprowadzamy w celu otrzymania wartości oznaczonej etykietą *wady_p*.

Przedstawione wnioskowanie przynosi dwie korzyści dla procesu identyfikacji polarności. Po pierwsze, w wielu przypadkach, w wyliczonym w opisany sposób wektorze cech dwie pierwsze cechy (*zalety* i *wady*) mogą mieć wartość zerową. Na rysunku 4.4 miałyby to miejsce wówczas, gdyby zarówno w recenzji 1 produktu A, jak i recenzji 3 produktu B recenzenci nie ocenili



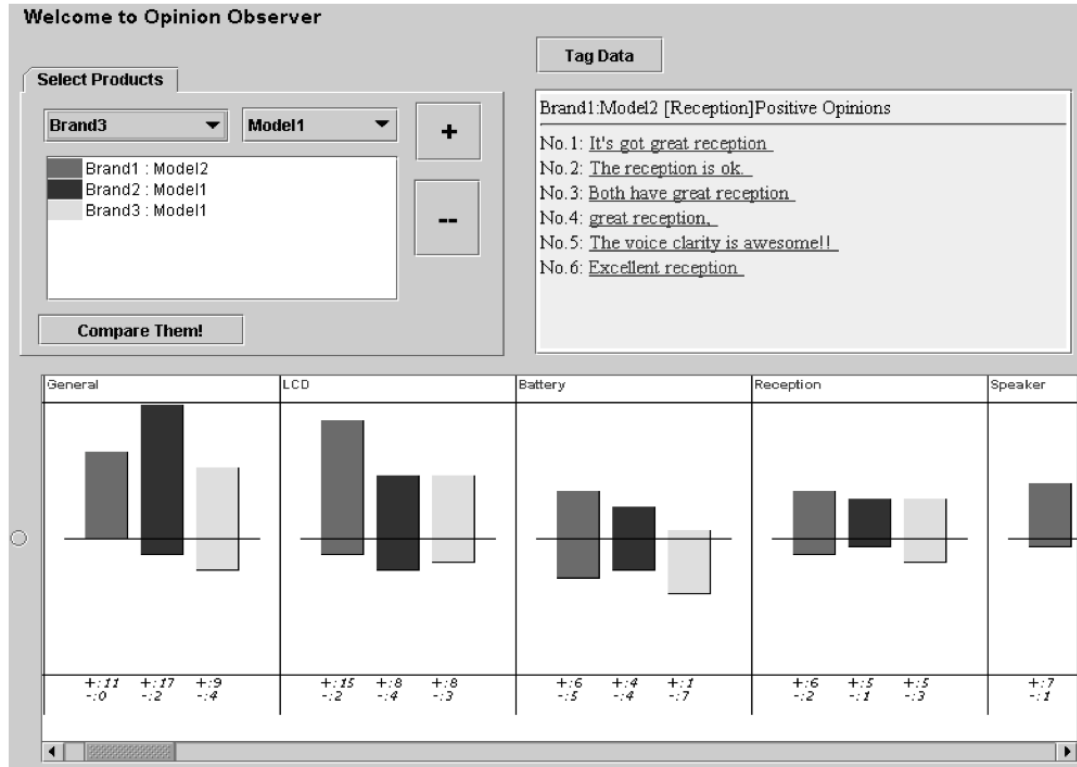
Rysunek 4.4: Proces generowania cech w proponowanej metodzie (wariant II). Źródło: opracowanie własne

aspektu *design* na liście zalet i wad. Jak zostanie wykazane w sekcji 5.3, taka sytuacja może mieć miejsce przede wszystkim wtedy, gdy analizowana para była użyta w niewielu recenzjach. W takim przypadku, wykorzystując wyłącznie dwie pierwsze cechy z wektora, nie byłoby możliwe przeprowadzenie identyfikacji polarności. Warto również zwrócić uwagę, że nawet wówczas, gdy dwie pierwsze cechy nie są zerowe, dodanie do wektora dwóch dodatkowych cech może zwiększyć poprawność przypisywania polarności, przede wszystkim w sytuacjach, gdy obie statystyki mają podobne wartości.

4.3.3 Wariant III

Kolejny sposób, w jaki potencjalnie możliwe byłoby wykorzystanie zaproponowanego wnioskowania przez systemy sztucznej inteligencji, jest o tyle ciekawy, że z założenia w ogóle nie wymaga dysponowania opiniami wyrażanymi przez recenzentów w przetwarzalny sposób. Sytuacja taka mogłaby mieć miejsce wówczas, gdy pewien system posiada już znaczną wiedzę z zakresu analizy wydźwięku i jest w stanie przygotować podsumowania opinii o produktach i usługach na podstawie wielu opinii wyrażonych o nich i ich aspektach w postaci wypowiedzi tekstowych. Istnieje wiele prac, których celem jest generowanie strukturyzowanych podsumowań opinii na podstawie wypowiedzi tekstowych. Przykładowo, opisany w artykułach [Hu i Liu 2004a; Liu, Hu i Cheng 2005] system OpinionObserver miał na celu analizę wielu recenzji dotyczących określonych produktów w celu wizualizacji podsumowań opinii recenzentów o różnych aspektach tych produktów.

Proponowane wnioskowanie mogłoby zostać wykorzystane w systemach o podobnych funkcjach co OpinionObserver w sytuacjach, gdy ten napotyka nowe, nieznanne sformułowania służące



Rysunek 4.5: Zrzut ekranu z systemu OpinionObserver, obrazujący podsumowania opinii wielu recenzentów o różnych aspektach trzech ocenianych telefonów komórkowych. Kolejne wykresy odnoszą się do różnych aspektów, natomiast na każdym wykresie słupki o odpowiednim kolorze reprezentują opinię recenzentów o konkretnym aspekcie dla danego produktu. Podsumowanie takie może zostać wygenerowane na podstawie analizy recenzji wyrażonych w postaci wypowiedzi tekstowych. Źródło: [Hu i Liu 2004a; Liu, Hu i Cheng 2005]

do wyrażania opinii o aspektach produktów i usług. System taki, dzięki opisanemu wnioskowaniu, może uczyć się polarności takich nowych słów. W sytuacji, gdy system znajdzie nieznaną słowo użyte do wyrażenia opinii o aspekcie dobra, możliwe jest, że system będzie posiadał wygenerowane podsumowanie opinii o tym produkcie czy usłudze, analogiczne do tego, którego wizualizacja jest zaprezentowana na rysunku 4.5. Podsumowanie takie byłoby wykorzystywane tak samo, jak miało to miejsce we wcześniej opisanych wariantach.

4.4 Założenia metody

Wnioskowania, przedstawione w opisanych wariantach proponowanej metody, wymagają spełnienia przez posiadany korpus recenzji pewnych założeń:

- recenzowane produkty i usługi są podzielone na ściśle określone kategorie, a każdy produkt (lub każda usługa) przypisany jest do jednej kategorii (np. „telefony komórkowe”, „pralki” czy „hotele”). Sytuacja taka jest często spotykana na portalach umożliwiających publikowanie recenzji konsumenckich, np. na portalu `cokupic.pl`. W przypadku przypisania produktu lub usługi do więcej niż jednej kategorii, fakt ten ponownie można postrzegać jako przypisanie do jednej, sztucznie utworzonej kategorii, obejmującej tylko te produkty i usługi, które były przypisane do dokładnie tych samych wielu kategorii,
- dla każdej kategorii produktów istnieje zasób wiedzy o aspektach, jakie mogą być poddawane ocenie przez recenzentów, możliwy do wykorzystania dla potrzeb automatycznego wnioskowania przez system informatyczny. Wiedzę taką rozumiemy jako płaską (czyli pozbawioną hierarchii) listę aspektów, gdzie do każdego aspektu przypisany jest leksykon sformułowań, za pomocą których recenzenci mogą się do niego odnosić (czyli nazw tego aspektu, np. synonimów odnoszących się do tego samego aspektu),
- produkty i usługi są jednoznacznie identyfikowalne lub przynajmniej są pogrupowane w klasy produktów i usług nierozróżnialnych z punktu widzenia recenzentów (przykładowo produkty o takiej samej nazwie marketingowej, różniące się od siebie w niewielkim stopniu konfiguracją parametrów technicznych). Dla każdej recenzji jesteśmy w stanie stwierdzić, jaki produkt lub usługa (lub jakie grupy produktów lub usług) podlega w niej ocenie. Sytuacja, w której dana recenzja jest przypisana do konkretnego, jednoznacznie identyfikowalnego produktu lub usługi, ma miejsce na wielu portalach umożliwiających publikowanie recenzji konsumenckich, np. na portalu `gastronauta.pl`. Niewątpliwie istnieją jednak sytuacje, gdy to założenie nie będzie spełnione. Przykładowo, pomimo tego, że recenzje dotyczyć będą produktów opisanych jedną nazwą marketingową, to produkty te będą się znacznie różniły. W związku z tym, konieczne jest dysponowanie metodą pozwalającą na wykrywanie takich sytuacji. Metoda analizy spójności opinii wyrażanych o produktach, opisana w rozdziale 5, pozwala na osiągnięcie m.in. tego celu,
- posiadane recenzje nie są recenzjami porównawczymi, to jest w każdej recenzji oceniany jest

tylko jeden produkt lub usługa,

- oprócz recenzji w postaci wypowiedzi tekstowych, dysponujemy przetwarzalnymi opiniami o produktach (np. ocenami punktowymi lub listami zalet i wad z predefiniowanymi opcjami), które można podsumować, aby poznać zbiorczą opinię wielu recenzentów o danym produkcie lub danej usłudze. Recenzje zawarte w obu wspomnianych kolekcjach nie muszą pochodzić od tych samych autorów. Inną możliwością jest to, że posiadamy odpowiednią metodę umożliwiającą generowanie takich podsumowań na podstawie przeanalizowanych dotychczas wypowiedzi tekstowych.

4.5 Podsumowanie

W rozdziale zaprezentowano obrane w pracy podejście do identyfikacji polarności sformułowań używanych do wyrażania opinii o aspektach produktów i usług. Podejście to polega na wykorzystaniu podsumowań opinii wielu recenzentów o aspektach, skąd czerpie się wskazówki dotyczące prawdopodobnej polarności słów, za pomocą których recenzenci oceniają poszczególne aspekty. Dzięki temu możliwe jest przeprowadzenie nauki polarności również wówczas, gdy nie posiadamy korpusu wymaganego w tradycyjnych podejściach nadzorowania odległego lub takie korpusy są niekompletne.

Zaproponowano trzy różne warianty, prezentujące różne sposoby, na jakie proponowane podejście mogłoby zostać wykorzystane do osiągnięcia zamierzonego celu. Warianty te różnią się stopniem skomplikowania wynikającego z konieczności pozyskania dodatkowych danych dotyczących podsumowań wielu opinii o produktach i usługach. Należy zaznaczyć, że zaproponowane podejście można wzbogacić o dalsze warianty, możliwe do zastosowania w różnych scenariuszach i dla różnych dostępnych danych.

Rozdział 5

Analiza spójności różnych sposobów wyrażania opinii o produktach i usługach

W rozdziale 4 zaprezentowano koncepcję wykorzystania podsumowań wielu opinii o produktach i usługach do nauki polarności sformułowań wykorzystywanych przez recenzentów do wyrażania opinii o różnych aspektach tych produktów i usług. Przed przystąpieniem do szczegółowego sformułowania sposobu działania proponowanej metody, przeprowadzono analizę danych. Jej celem było zbadanie zgromadzonych danych oraz wstępne potwierdzenie, iż proponowane wnioskowania mogą być wykorzystane w procesie nauki polarności.

Obszerne fragmenty tego rozdziału opublikowano w artykule [Małyszko 2013]. Przeprowadzono również dodatkowe eksperymenty, opublikowane w artykułach [Bernatowicz i Małyszko 2014; Małyszko i Filipowska 2013], które nie zostały opisane w prezentowanej pracy ze względu na zapewnienie jej odpowiedniej zwięzłości.

5.1 Motywacja badań

W sekcji 4.1 zwrócono uwagę na fakt, że aby przeprowadzić wnioskowanie zgodne z koncepcją nadzorowania odległego, konieczne jest spełnienie wymagających założeń. W takich metodach niezbędne jest przede wszystkim posiadanie opinii, wyrażonych przez tych samych recenzentów na różne sposoby, tj. w postaci niestrukturyzowanej (np. w postaci wypowiedzi tekstowych) oraz

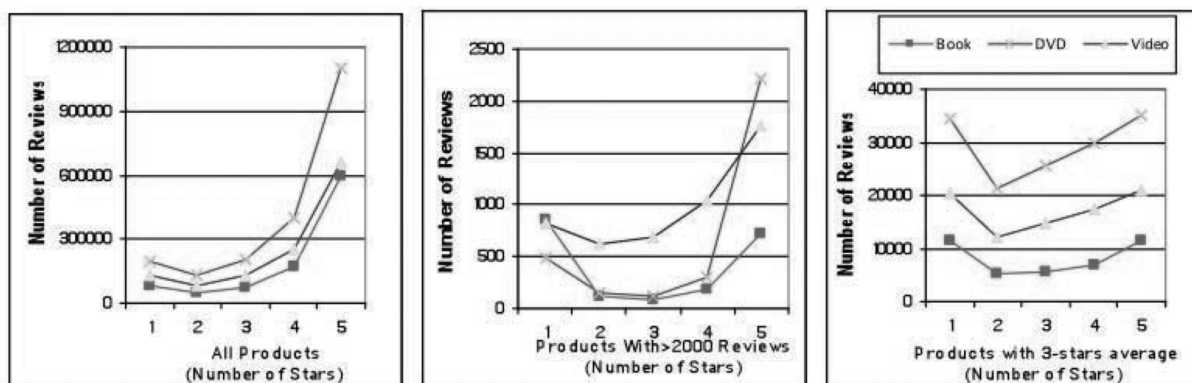
strukturyzowanej (np. w postaci ocen punktowych). Istniejące metody z nurtu nadzorowania odległego nie mogą być zastosowane w sytuacji, gdy, choć posiadamy dwie kolekcje recenzji dotyczących tych samych produktów lub usług, opinie w nich zawarte wyrażane są przez różnych recenzentów. Podobnie, nadzorowanie odległe nie może być przeprowadzone w sytuacji, gdy strukturyzowana i niestrukturyzowana kolekcja są względem siebie niekompletne, tj. w jednej z kolekcji znajduje się informacja, która nie znajduje odzwierciedlenia w kolekcji drugiej. Pierwszym celem tego rozdziału jest wykazanie, że istnieją korpusy, w których faktycznie wspomniane niekompletności danych występują.

Dodatkowo, celem rozdziału jest dokładniejsze przeanalizowanie podsumowań wielu opinii o produktach i usługach. Istotnym wymaganiem, stawianym przez proponowaną metodę wykorzystywanym korpusom danych, jest spójność opinii, wyrażanych przez różnych użytkowników na temat tych samych dóbr zawartych we wspomnianych korpusach. Przez spójny sposób postrzegania produktów i usług rozumiemy sytuację, gdzie wśród ocen nadawanych przez recenzentów poszczególnym produktom i usługom oraz ich aspektom powinno być możliwe wykazanie tendencji. Przykładem takiej tendencji jest sytuacja, w której pewien aspekt w określonym produkcie prawie zawsze jest oceniany pozytywnie, natomiast inny aspekt raczej negatywnie. Oczywiście, nie oznacza to, że wszyscy recenzenci muszą mieć takie same opinie o produktach, usługach i ich aspektach, jednak w przypadku przeanalizowania wielu opinii o danym produkcie czy usłudze taka tendencja musi występować. W przypadku braku jej występowania, podsumowania wielu opinii o produkcie lub usłudze tworzyłyby szum, na podstawie którego wnioskowania opisane w rozdziale 4 nie przyniosłyby prawidłowych wyników. Przykładowo, sytuacja taka może wystąpić wówczas, gdy recenzje dotyczą różnych produktów, które błędnie zostały rozpoznane jako ten sam produkt (np. były to produkty o podobnych nazwach, ale bardzo różniące się specyfikacją techniczną).

Jak wspomniano w rozdziale 2, podsumowania wielu opinii o produktach i usługach często są wykorzystywane w celu umożliwienia użytkownikom szybkiego zapoznania się z tym, w jaki sposób dany produkt lub usługa są oceniane. Przykłady takich podsumowań zaprezentowano na rysunkach 2.2 na stronie 25 oraz 2.3 na stronie 26. Należy jednak zwrócić uwagę na fakt, że istnieją publikacje, w których wykazano, że rozkłady ocen nadawanych przez recenzentów danemu dobru posiadają często pewne cechy utrudniające wnioskowanie na ich podstawie.

Przykładowo, w pracy [Chevalier i Mayzlin 2003] zaobserwowano, że zdecydowana większość ocen recenzentów w analizowanym korpusie była pozytywna (średnia ocen przekraczała 4 punk-

ty w skali pięciopunktowej). W pracach [Hu, Pavlou i Zhang 2007; Hu, Zhang i Pavlou 2009] z powodu takiej obserwacji stwierdzono, że średnia ocen, standardowo wykorzystywana w podsumowaniach opinii wyrażonych w postaci punktowej, w wielu przypadkach nie powinna być w tym celu stosowana. Powodem jest fakt, że średnia arytmetyczna posiada statystyczny sens tylko wówczas, gdy rozkład uśrednianych wartości jest unimodalny lub symetrycznie bimodalny, podczas gdy faktycznego rozkładu ocen często nie można zaliczyć do żadnej z tych opcji. Z badań przeprowadzonych w pracy [Hu, Zhang i Pavlou 2009] wynika, że rozkłady ocen punktowych nadawanych produktom mają z reguły rozkład J-kształtny (ze znaczną przewagą ocen pozytywnych), jak zaprezentowano na rysunku 5.1. Prawdopodobnie z tego właśnie powodu, liczne badania wykorzystujące podsumowania opinii o produktach (przykładowo do prognozowania sprzedaży) osiągały niekonkluzywne wyniki (w różnych artykułach zajmujących się tymi samymi zagadnieniami otrzymywano sprzeczne rezultaty) [Hu, Pavlou i Zhang 2007; Hu, Zhang i Pavlou 2009] (patrz sekcja 2.5.1).



Rysunek 5.1: Rozkłady wszystkich ocen punktowych nadanych różnym kategoriom produktów (od lewej: wszystkim produktom, produktom z liczbą recenzji przekraczającą 2000, ze średnią ocen na poziomie 3). Źródło: [Hu, Zhang i Pavlou 2009]

W związku z zaistnieniem opisywanych zjawisk rodzi się pytanie czy proponowane w pracy podejście faktycznie może być wykorzystane w procesie nauki polarności. Analiza tego zagadnienia, wraz z analizą działania metod z nurtu nadzorowania odległego w sytuacji niekompletnych danych, stanowią motywację badań zaprezentowanych w tym rozdziale pracy. W eksperymentach zaprezentowanych poniżej przeprowadzono statystyczną weryfikację tego, czy sposób oceniania określonych produktów i usług w posiadanym korpusie umożliwia stosowanie proponowanego podejścia. W ten sposób zrealizowany ma zostać pierwszy cel szczegółowy rozprawy (patrz strona 9).

5.2 Koncepcja eksperymentów

W eksperymentach zaprezentowanych w rozdziale posłużono się danymi opisanymi w sekcji 4.3.2, możliwymi do zastosowania w drugim wariantcie proponowanej metody. W sekcji opisano metodę nauki polarności sformułowań używanych do wyrażania opinii o różnych aspektach telefonów komórkowych z wykorzystaniem korpusu pobranego z portalu `cokupic.pl`. W podsekcjach poniżej opisano przyjęte cele poszczególnych eksperymentów oraz omówiono wykorzystane dane.

Należy zaznaczyć, że eksperymenty opisane w rozdziale były przeprowadzone na danych, których nie można traktować jako próbki reprezentującej całą populację recenzji konsumenckich publikowanych w Internecie. Opisane eksperymenty miały wykazać, że istnieją przypadki, dla których zaproponowana metoda może poprawić wyniki identyfikacji polarności.

5.2.1 Cele eksperymentów

Zaplanowano trzy eksperymenty, mające za zadanie analizę przykładowych, pozyskanych przez autora z Internetu, danych dla przetestowania proponowanego podejścia. Celem eksperymentów było formalne wykazanie, że proponowana metoda może pozwolić na uzyskanie oczekiwanych rezultatów. Przeprowadzono następujące eksperymenty:

1. Eksperyment mający na celu określenie podobieństwa informacji zawartej w wypowiedzi tekstowej do tej zamieszczonej na liście zalet i wad w ramach pojedynczych recenzji. Istotą eksperymentu było sprawdzenie, w jakim stopniu nadane przez recenzentów zalety i wady mogą służyć jako wskazówki w procesie identyfikacji polarności (na zasadach zgodnych z ideą nadzorowania odległego). Dodatkowo przeanalizowano, czy występują przypadki sprzecznych informacji występujących w obu częściach recenzji. Eksperyment opisano w sekcji 5.3.
2. Eksperyment, w trakcie którego badano hipotezę, że produkty i usługi są postrzegane przez recenzentów w pewien spójny, określony sposób. Miał on wykluczyć możliwość, że opinie recenzentów o produktach są na tyle niespójne, że nie jest możliwe przeprowadzanie wnioskowań na podsumowaniach wielu recenzji pojedynczego produktu czy usługi. Eksperyment polegał na badaniu tego, czy podsumowania opinii o tym samym produkcie, ale wygenerowane na podstawie różnych jego recenzji, są do siebie podobne. Eksperyment opisano w sekcji 5.4.

3. Eksperyment mający na celu określenie, na ile podsumowania opinii o tym samym produkcie, wyrażone w postaci wypowiedzi tekstowych oraz w postaci list zalet i wad, są do siebie podobne. Eksperyment opisano w sekcji 5.5.

5.2.2 Posiadane dane i procedura ich przygotowania

Eksperymenty przeprowadzono z wykorzystaniem korpusu recenzji telefonów komórkowych, pobranego ze strony `cokupic.pl` w dniu 1 lutego 2013. Korpus składał się m.in. z recenzji wyrażonych za pomocą wypowiedzi tekstowej (tekstu w języku naturalnym) i list zalet i wad. Pełen korpus liczył 77.936 recenzji, które dotyczyły 1858 produktów. Średnio jeden produkt posiadał 41,946 recenzji, zaś kwartylami liczby recenzji nadanych poszczególnym produktom są:

- pierwszy kwartył – 4 recenzje,
- drugi kwartył (mediana) – 11 recenzji,
- trzeci kwartył – 34 recenzji,
- czwarty kwartył (najwyższa wartość) – 2552 recenzji.

W posiadanych recenzjach, produkty są oceniane jednocześnie za pomocą tekstu w języku naturalnym, jak i listy zalet i wad. Możliwe zalety i wady są ustalone dla danej kategorii, tj. użytkownik nie opisuje zalet i wad, lecz wybiera je z zaproponowanej mu listy, jak to zaprezentowano na rysunku 5.2. Dodatkowo, każdej zalecie odpowiada wada (zalety i wady dotyczą tych samych aspektów). Listy zalet i wad, przygotowane zgodnie z opisanym formatem, mogą stanowić strukturalizowaną kolekcję, możliwą do wykorzystania w przetwarzaniu wypowiedzi tekstowych zgodnie z założeniami nadzorowania odległego.

W korpusie produkty (telefony komórkowe) oceniane były za pomocą 23 możliwych zalet i wad. Każda z tych zalet i wad odnosiła się do pewnego aspektu produktu. Przed przystąpieniem do eksperymentu, na podstawie analizy list zalet i wad zdefiniowano 22 aspekty i do każdego z nich przypisano oceniające je zalety i wady (w jednym przypadku postanowiono przypisać do jednego aspektu dwie możliwe zalety i dwie możliwe wady ze względu na trudność rozgraniczenia znaczeniowego pomiędzy nimi; były to „łatwość obsługi” i „łatwość użycia”).

Część recenzji wyrażonych w postaci tekstu w języku naturalnym została ręcznie zaadnotowana. Były to recenzje losowej próby 16 modeli telefonów komórkowych (w sumie 216 recenzji).

Plusy *	Minusy *
<input type="checkbox"/> cena	<input type="checkbox"/> cena
<input type="checkbox"/> szybkość	<input type="checkbox"/> wolny
<input type="checkbox"/> dobry aparat fotograficzny	<input type="checkbox"/> kiepski aparat fotograficzny
<input type="checkbox"/> długa żywotność baterii	<input type="checkbox"/> słaba bateria
<input type="checkbox"/> lekkość	<input type="checkbox"/> ciężki
<input type="checkbox"/> wytrzymałość	<input type="checkbox"/> delikatna konstrukcja
<input type="checkbox"/> łatwość użycia	<input type="checkbox"/> trudność użycia

Rysunek 5.2: Zrzut ekranu z portalu cokupic.pl prezentujący fragmenty listy zalet i wad, za pomocą której recenzenci mogą wyrażać swoje opinie o produktach z kategorii „Telefony komórkowe”. Źródło: cokupic.pl. Dostęp: 30 kwietnia 2014

Produkty zostały wylosowane z populacji ograniczonej w ten sposób, że brano pod uwagę wyłącznie produkty o liczbie recenzji nie większej niż 34. Było to spowodowane tym, że istniało ryzyko wylosowania produktu o bardzo dużej liczbie recenzji (przykładowo, przekraczającej 100 recenzji), w wyniku czego proces adnotacji byłby bardzo czasochłonny. W związku z tym, podjęto decyzję o odrzuceniu z populacji jednej czwartej produktów z największą liczbą recenzji (jak wspomniano, 34 recenzje to trzeci kwartył rozkładu liczby produktów względem liczby recenzji). Dodatkowo, wprowadzono też ograniczenie dolne, tj. losowano produkty z co najmniej 5 recenzjami. Tak ograniczona populacja wynosiła 10.271 recenzji 731 produktów. Wnioskowania przeprowadzane na podstawie tak wylosowanych próbek będą dotyczyły tylko części populacji, z której przeprowadzono losowanie. Rozmiar próbki wybrano arbitralnie, biorąc pod uwagę przewidywaną czasochłonność zadania. Jej wielkość jest jednak porównywalna do próbek wykorzystywanych w podobnych badaniach z zakresu analizy języka naturalnego.

Adnotacja polegała na czytaniu przez adnotatora recenzji tekstowych i przypisywaniu do nich zalet i wad z zadanej puli na podstawie informacji zawartej tylko w tekście (lista zalet i wad nadana przez recenzenta była przed adnotatorem zakryta). Adnotacja była przeprowadzona zgodnie ze zdefiniowanymi wskazówkami adnotacji, opisanymi w załączniku A na stronie 177. Proces adnotacji został przeprowadzony dwukrotnie, tj. te same dokumenty były dwukrotnie adnotowane przez tego samego adnotatora w dużym odstępie czasu (tj. ponad miesiąc) pomiędzy obydwojema podejściami do adnotacji, w celu zapewnienia jak najbardziej obiektywnych adnotacji.

Ponieważ dysponowano dwoma zestawami adnotacji dla dokumentów, w których, z względu na bogactwo języka naturalnego i niejednoznaczność wypowiedzi mogły występować niejedno-

znacznosci, konieczne bylo ich polaczenie w celu utworzenia tzw. zlotego standardu, ktory bylyby uznany jako obowiazujacy zestaw adnotacji dla dokumentow. Podjeto decyzje o przygotowaniu dwuch takich zestawow:

- zestaw adnotacji, w ktorym dla kazdego dokumentu adnotacjami jest suma zbiorow adnotacji z obu laczonych zestawow,
- zestaw adnotacji, bedacy dla kazdego dokumentu iloczynem adnotacji dla tego dokumentu z obu laczonych zestawow.

LiczebnoSci przygotowanych w ten sposob danych zaprezentowano w tabeli 5.1.

Źródło	Liczba zalet	Liczba wad	Razem
listy zalet/wad	860	399	1259
adnotacje tekstowe (suma)	501	297	798
adnotacje tekstowe (iloczyn)	362	214	576

Tabela 5.1: Statystyki korpusu zaadnotowanych recenzji. Dwa dolne wiersze prezentuja statystyki dotyczace dwuch wariantow zlotego standardu adnotacji: bedacego wynikiem sumy lub iloczynu adnotacji z dwuch laczonych zestawow. Źródło: opracowanie własne

5.3 Eksperyment 1. Spójność informacji na poziomie pojedynczych recenzji

Pierwszy eksperyment przeprowadzono w celu okreslenia podobienstwa informacji zawartej w opisie tekstowym do tej zamieszczonej w liście zalet i wad w ramach pojedynczych recenzji. Mial on za zadanie sprawdzic, w jakim stopniu takie nadane przez recenzentow zalety i wady moga sluzyc jako wskazowki w procesie identyfikacji polarnosci sformulowan uzywanych w wypowiedziach tekstowych. Niskie podobienstwo pomiedzy informacjami zawartymi w obu czesciach pojedynczych recenzji wskazałyby, ze nadzorowanie odlegle z wykorzystaniem danych pochodzacych tylko z pojedynczych recenzji (to jest zarowno wypowiedz tekstowa, jak i lista zalet i wad musza pochodzic z tej samej recenzji) w wielu przypadkach nie pozwolilyby na prawidlowe rozpoznanie polarnosci. Tym samym, krytyka istniejacych metod nadzorowania odleglego opisana w sekcjach 4.1 i 4.3.2 zostalaby uzasadniona, co potwierdzilyby istotnosc przyjetego kierunku badan.

5.3.1 Koncepcja eksperymentu

W przeprowadzonym eksperymencie, kolejno analizowano pojedyncze recenzje i porównywano adnotacje nadane przez adnotatora wypowiedziom tekstowym (zgodnie z opisem w sekcji 5.2.2) z zaletami i wadami nadanymi w liście przez recenzenta. Ręczne adnotacje wykorzystano jako tzw. złoty standard, do którego porównywano zalety i wady nadane przez recenzentów, dzięki czemu możliwe było przeprowadzenie oceny zgodności informacji zawartych w obu kolekcjach. Zgodność ustalano dla obu zestawów przygotowanych adnotacji zaprezentowanych w tabeli 5.1 i omówionych w sekcji 5.2.2. Ocenę przeprowadzono z wykorzystaniem miar precyzji, pełności oraz F1. Wspomniane statystyki dla poszczególnych recenzji obliczano zgodnie z następującymi formułami:

$$\text{Precyzja} = \frac{\text{poprawne}}{\text{poprawne} + \text{nadmiarowe}} \quad (5.1)$$

$$\text{Pełność} = \frac{\text{poprawne}}{\text{poprawne} + \text{pominięte}} \quad (5.2)$$

$$F_{\beta} = \frac{(\beta^2 + 1)\text{Precyzja} * \text{Pełność}}{(\beta^2\text{Precyzja}) + \text{Pełność}} \quad (5.3)$$

gdzie:

- jako poprawne rozumiemy te informacje (adnotacje), które znalazły się w obu kolekcjach; w analizowanym przypadku są to adnotacje, które pojawiły się jednocześnie w obu częściach recenzji (jednocześnie w opisie tekstowym oraz w liście zalet i wad),
- nadmiarowe informacje to takie, które nie znalazły się w złotym standardzie, a które pojawiły się w porównywanej do niego kolekcji; w naszym przypadku są to adnotacje, które w danej recenzji pojawiły się wyłącznie w liście zalet i wad (nie nadano odpowiedniej adnotacji w recenzji tekstowej),
- mianem pominiętych określamy informacje, które znajdują się w złotym standardzie i których brakuje w porównywanej kolekcji; u nas są to adnotacje, które pojawiły się wyłącznie w recenzji tekstowej, a recenzent nie zawarł ich w liście zalet i wad,
- parametr β różnicuje wagę precyzji i pełności przy obliczaniu statystyki F_{β} . Przy parametrze β na poziomie 1, obie statystyki mają jednakowy wpływ na wartość statystyki F_{β} .

Precyzja ukazuje więc, na ile pozycje nadane przez recenzentów na listach zalet i wad odpowiadają adnotacjom z recenzji tekstowej, natomiast pełność mówi o tym, w jakim stopniu adnotacje z recenzji tekstowych znajdują odzwierciedlenie w pozycjach z list zalet i wad. Statystykę F_β interpretuje się w charakterze harmonicznej średniej ważonej powyższych statystyk. Wyniki uzyskane dla poszczególnych recenzji zostały następnie uśrednione dla całej analizowanej próbki.

Dodatkowo, w eksperymencie analizowano recenzje pod kątem występowania w nich sprzecznych informacji. Działania opisane w powyższych akapitach sprawdzały jedynie czy informacje wynikające z recenzji tekstowych oraz list zalet i wad były takie same, nie sprawdzały jednak czy miały miejsce przypadki sprzeczności pomiędzy obiema częściami recenzji. Przykładowo, w części tekstowej mogła się pojawić adnotacja oznaczająca, że recenzent ocenił pewien aspekt pozytywnie, natomiast jednocześnie recenzent oznaczył ten sam aspekt na liście wad. Należało więc dodatkowo przeanalizować, jak często opisana sytuacja ma miejsce w posiadanym korpusie.

5.3.2 Procedura eksperymentu i uzyskane wyniki

Omawiany eksperyment przeprowadzono z wykorzystaniem skryptu przygotowanego w języku Python, który dla kolejnych recenzji pobierał z bazy danych nadane przez recenzentów zalety i wady oraz adnotacje nadane przez adnotatora, a następnie dokonywał porównania, na podstawie którego wyliczał odpowiednie statystyki. Eksperyment przeprowadzono dla danych, których statystyki zaprezentowano w tabeli 5.1.

Złoty standard	Precyzja	Pełność	F_1	Sprz.
adnotacje tekstowe (suma)	0,239	0,465	0,281	14
adnotacje tekstowe (iloczyn)	0,286	0,435	0,312	5

Tabela 5.2: Statystyki zgodności list zalet i wad z adnotacjami tekstowymi pochodzącymi ze złotych standardów. W kolumnie „Sprz.” (sprzeczność) znajduje się liczba przypadków w całym korpusie, w którym informacje z recenzji tekstowych oraz z list zalet i wad o pewnym aspekcie były sprzeczne. Źródło: opracowanie własne

Otrzymane wyniki zaprezentowano w tabeli 5.2. W przypadku 38 recenzji pominięto obliczanie statystyk precyzji i pełności z powodu braku jakichkolwiek adnotacji w recenzji tekstowej, bądź z zalet i wad w liście (gdyż w tym przypadku powstaje problem dzielenia przez zero przy obliczaniu

tych statystyk).

5.3.3 Interpretacja wyników

Uzyskane wyniki wskazują, że w przypadku pojedynczej recenzji informacje wyrażone w różnych częściach recenzji (wyrażonych na różny sposób) w znacznym stopniu różnią się między sobą. Lepsze rezultaty otrzymano dla złotego standardu będącego wynikiem iloczynu zbiorów adnotacji, na nim więc skoncentrujemy się w dalszej analizie.

Uzyskane statystyki można zinterpretować w następujący sposób:

- spośród opinii wyrażonych w tekście tylko średnio około 28,6% opinii znajduje odzwierciedlenie na liście zalet i wad przypisanych do tej samej recenzji, czyli dla ponad 70% opinii nie istnieje na liście zalet i wad informacja, którą można by wykorzystać do określania polarności sformułowań wykorzystanych do wyrażania tych opinii (zgodnie z podejściem nadzorowania odległego),
- tylko 43% opinii wyrażonych w postaci list zalet i wad znajduje odwzorowanie w wypowiedziach tekstowych przygotowanych przez tego samego recenzenta, co oznacza, że prawie 60% takich opinii z list zalet i wad nie może być wykorzystane w opisywanym podejściu w procesie nauki polarności.

Należy zwrócić uwagę na fakt, że pomimo uzyskania niskiego podobieństwa pomiędzy informacjami pochodzącymi ze wspomnianych części recenzji, to z reguły nie są to informacje sprzeczne. Przy ogólnej liczbie adnotacji nadanych wypowiedziom tekstowym na poziomie 576 znaleziono jedynie 5 sprzeczności, co oznacza, że przeciwna polarność została przypisana do sformułowań w tekście w mniej niż 1% przypadków.

W celu wykrycia przyczyny występowania sprzecznych informacji przeanalizowano wszystkie przykłady sprzeczności dla obu złotych standardów. Wykryto różne powody, dla których sprzeczności takie występowały. W niektórych przypadkach były to nieścisłości w ocenianiu; przykładowo, w jednej recenzji w opisie tekstowym znalazł się fragment „jedyny minus to brak normalnej wibracji i zasięg w niektórych odludnych miejscach nie łapie” (pisownia oryginalna), natomiast w liście zalet i wad, z nieznanymi przyczynami, recenzent oznaczył sygnał jako zaletę. Powodem sprzeczności była też sytuacja, gdy pewne aspekty jednocześnie w pewnym stopniu były postrzegane przez recenzenta i jako zalety, i jako wady. Na przykład, na podstawie fragmentu recenzji telefonu komórkowego, w którym oceniany był wbudowany aparat fotograficzny, „Świetny aparat,

bo 8 Mpx, który ma jednak jedną podstawową wadę - nie posiada zoomu!!!” adnotator nadał adnotację, oznaczającą aparat fotograficzny jako zaletę, podczas gdy recenzent w liście zalet i wad oznaczył ten aspekt jako wadę.

Natknęto się również na przypadek, w którym część tekstowa jednej z recenzji, w których wykryto sprzeczność, była w istocie w całości przekopiowanym, krążącym po wielu portalach tekstem, za pomocą którego oceniano co najmniej 7 typów telefonów komórkowych (fakt ten wykryto z wykorzystaniem wyszukiwarki internetowej). W takich recenzjach, pomimo dokładanie tego samego tekstu, oznaczano na listach różne zalety i wady, co powodowało wystąpienie sprzeczności. Prawdopodobnie należy uznać taką recenzję za przykład omawianego w sekcji 2.5.2 zjawiska *opinion spam* (na co zwrócono uwagę m.in. w pracy [Jindal i Liu 2007]). Takie recenzje są oczywiście niepożądane i zaburzają uzyskiwane wyniki, jednak ich wykrywanie i eliminowanie z korpusu jest poza zakresem prezentowanej pracy.

Jeszcze jednym możliwym powodem występowania sprzeczności w ramach pojedynczych recenzji jest to, że czasem recenzenci postrzegali produkt bardzo negatywnie, w związku z czym na liście zalet i wad zaznaczali (prawdopodobnie bez zastanowienia) wszystkie możliwe wady pomimo tego, że w recenzji tekstowej wybrany aspekt oceniali jako zaletę.

Podsumowując, w analizowanym korpusie, w przypadku nauki polarności zgodnie z koncepcją nadzorowania odległego na poziomie pojedynczych recenzji należy oczekiwać, że ponad połowa informacji z list zalet i wad nie może być wykorzystana do tego celu. W obu kolekcjach (wypowiedzi tekstowych oraz listach zalet i wad) recenzenci często zamieszczają inne informacje. Jednocześnie, na podstawie niewielkiej liczby wykrytych sprzeczności, z wysoką pewnością możemy przypuszczać, że jeśli jakiś aspekt jest oceniany w obu częściach recenzji, to pozycja z listy zalet i wad trafnie wskazuje na to, jaką polarność należy przypisać sformułowaniu użytemu do oceny tego aspektu w wypowiedzi tekstowej. Potwierdza to obserwację zawartą w sekcji 4.1 i ukazuje, że założenia przyjęte w dotychczasowych podejściach z nurtu nadzorowania odległego w pewnych przypadkach mogą okazać się niespełnione – posiadane kolekcje, które mają zostać wykorzystane do nadzorowania odległego, mogą posiadać różne informacje, co sprawia, że traktowanie jednej kolekcji (strukturyzowana kolekcja nie opisuje w pełni kolekcji drugiej – niestrukturyzowanej).

5.4 Eksperyment 2. Badanie spójności opinii wielu recenzentów

W tej sekcji zaprezentowany zostanie eksperyment, mający na celu wykazanie, że produkty są postrzegane przez recenzentów w pewien spójny, określony sposób. Innymi słowy, analizowano, czy da się określić pewną tendencję w ocenianiu poszczególnych produktów i ich aspektów. Jeśli taka tendencja występuje, to znając podsumowanie opinii wielu recenzentów o danym bycie, możemy z pewnym prawdopodobieństwem przewidzieć, jaką opinię o tym produkcie może wyrazić kolejny recenzent, a tym samym jaka jest polarność sformułowania, którego użyje do oceny tego produktu lub usługi (przy założeniu, że wyraża opinię w postaci wypowiedzi tekstowej).

Koncepcja eksperymentu

W omawianym eksperymencie analizowano wyłącznie opinie wyrażone w postaci list zalet i wad z predefiniowanymi opcjami i określano czy możliwe jest wykrycie tendencji w ocenach nadawanych w ten sposób poszczególnym produktom. Opinie wyrażane w tym formacie są bardzo łatwe do przetwarzania, w związku z czym możliwe było przeprowadzenie eksperymentu na bardzo dużej ilości danych, bez konieczności przeprowadzania dodatkowych etapów przygotowujących dane (takich jak adnotacja wypowiedzi tekstowych).

Przed przeprowadzeniem eksperymentu przyjęto następujące założenia:

- możliwe jest określenie odległości pomiędzy podsumowaniami opinii wielu recenzentów o produktach,
- podobieństwo pomiędzy dwoma podsumowaniami opinii o tym samym produkcie, ale wygenerowanymi na podstawie rozłącznych kolekcji recenzji, powinno być większe, niż podobieństwo pomiędzy podsumowaniami dwóch różnych produktów. Oznacza to, że możliwe jest wykazanie tego, że średnia odległość pomiędzy dwoma podsumowaniami tych samych produktów jest istotnie mniejsza od średniej odległości pomiędzy dwoma podsumowaniami produktów różnych. Oczywiście, nie zawsze taka odległość dla podsumowań tego samego produktu będzie niższa niż dla produktów różnych. Wśród wszystkich produktów ocenianych w korpusie z pewnością będą występowały bardzo podobnie oceniane produkty, dla których podsumowania będą bardzo podobne. Statystycznie jednak rzecz biorąc, podobieństwo pomiędzy podsumowaniami opinii o tym samym produkcie powinno być większe niż pomiędzy podsumowaniami dwóch różnych, losowych produktów.

Dane do eksperymentu przygotowano w ten sposób, że zbiory recenzji wyrażonych w postaci list zalet i wad dla każdego produktu podzielono na dwie równe części (w przypadku nieparzystej liczby recenzji danego produktu, liczebności obu części różniły się o jeden). Dla każdego takiego podzbioru recenzji danego produktu, wygenerowano podsumowanie opinii wyrażonych w recenzjach do niego należących. W ten sposób, posiadano po dwa podsumowania opinii dla każdego wybranego produktu, wygenerowane na podstawie rozłącznych zbiorów recenzji tego produktu. Przykład takich dwóch podsumowań wygenerowanych dla pojedynczego produktu zaprezentowano w tabeli 5.3. Kolejne wiersze tabeli odpowiadają wybranym aspektom produktu.

Aspekt	Podsumowanie 1		Podsumowanie 2	
	Zalet	Wad	Zalet	Wad
waga	99	3	80	4
wytrzymałość	12	71	22	56
aparat fotograficzny	82	10	79	10
rozmiar wyświetlacza	71	1	58	0
rozdzielczość wyświetlacza	44	5	37	1
design	65	7	54	4
awaryjność	0	45	1	37

Tabela 5.3: Fragmenty podsumowań wygenerowanych na podstawie dwóch rozłącznych zbiorów recenzji telefonu komórkowego Samsung U700. Źródło: opracowanie własne na podstawie strony <http://cokupic.pl/produkt/Samsung-U700>. Dostęp: 1 lutego 2013

Jak wspomniano, aby wnioskowanie zaproponowane w rozdziale 4 było uzasadnione, powinno być możliwe wypracowanie miary, zgodnie z którą podsumowania opinii o jednym, tym samym produkcie (ale wygenerowane na podstawie innych jego recenzji) będą do siebie bardziej podobne, niż podsumowania opinii o dwóch różnych produktach.

Przykładowo, w tabeli 5.3 widać pewne podobieństwo pomiędzy dwoma podsumowaniami tego samego produktu (telefonu Samsung-U700). Analiza obu podsumowań wskazuje, że dla tych samych aspektów w obu kolekcjach wyrażane opinie były do siebie w pewien sposób podobne. Przykładowo, do wagi i designu recenzenci byli zdecydowanie pozytywnie nastawieni, podczas gdy wytrzymałość i awaryjność ocenili negatywnie. Eksperyment miał na celu w formalny sposób zweryfikować czy faktycznie w korpusie da się stwierdzić zależność, zgodnie z którą takie

	s_{1Y}	s_{2Y}	...	s_{nY}
s_{1X}	$d(s_{1X}, s_{1Y})$	$d(s_{1X}, s_{2Y})$...	$d(s_{1X}, s_{nY})$
s_{2X}	$d(s_{2X}, s_{1Y})$	$d(s_{2X}, s_{2Y})$...	$d(s_{2X}, s_{nY})$
...
s_{nX}	$d(s_{nX}, s_{1Y})$	$d(s_{nX}, s_{2Y})$...	$d(s_{nX}, s_{nY})$

Tabela 5.4: Tablica odległości dla wszystkich par z iloczynu kartezjańskiego $X \times Y$ pomiędzy wszystkimi podsumowaniami wygenerowanymi na podstawie recenzji należących do zbiorów X i Y . Źródło: opracowanie własne

podobieństwa podsumowań tych samych produktów są wysokie.

5.4.1 Kryterium spójności postrzegania produktów przez recenzentów

Weryfikację tego, czy faktycznie w posiadanym korpusie występuje opisywana zależność, przeprowadzono w następujący sposób. Podzielono posiadane podsumowania produktów na dwie części - X i Y , gdzie w każdej części znajdowało się dokładnie jedno podsumowanie każdego z analizowanych produktów (jak wspomniano, dla każdego produktu wygenerowano po dwa podsumowania, każde z nich na podstawie połowy jego dostępnych recenzji). Następnie, obliczono podobieństwa poszczególnych podsumowań należących do zbioru X do wszystkich podsumowań należących do zbioru Y . Innymi słowy, obliczono podobieństwa dla wszystkich par należących do iloczynu kartezjańskiego zbiorów X i Y . W wyniku powstała tablica odległości o strukturze zaprezentowanej w tabeli 5.4. Oznaczenia w tej tabeli mają następującą interpretację:

- s_{iX} to podsumowanie recenzji i -tego produktu pochodzące ze zbioru X , zaś s_{jY} produktu j -tego pochodzące ze zbioru Y ,
- $d(s_{iX}, s_{jY})$ to odległość pomiędzy s_{iX} a s_{jY} .

Dla tak wygenerowanej tablicy możliwe było określenie tego, jak się mają wartości $d(s_{iX}, s_{jY})$ dla $i = j$ (czyli podobieństwa podsumowań tych samych produktów) w porównaniu z pozostałymi przypadkami, tj. gdy obliczana jest odległość pomiędzy podsumowaniami różnych produktów. W związku z tym, jeśli opinie o produktach wielu recenzentów są spójne, wartości dla $i = j$ (czyli wartości na przekątnej w tabeli 5.4) powinny być systematycznie niższe od pozostałych wartości w tabeli.

Aby to sprawdzić, przeprowadzono następujące testy statystycznie (jeden parametryczny i jeden nieparametryczny):

1. wyliczono średnią wartość dla $d(s_{iX}, s_{jY})$ dla $i = j$ oraz dla pozostałych przypadków, a następnie przeprowadzono statystyczny test istotności dla średnich tych dwóch populacji,
2. pomiędzy wspomnianymi populacjami przeprowadzono test Manna-Whitneya na zgodność rozkładów z punktu widzenia przyjmowanych wartości odległości $d(s_{iX}, s_{jY})$ w analizowanych obserwacjach.

5.4.2 Miary odległości między podsumowaniami

W celu przeprowadzenia operacji opisanych w poprzedniej sekcji, konieczne jest wypracowanie miary odległości pomiędzy podsumowaniami. W trakcie prac przygotowano kilka miar, za pomocą których porównywano podobieństwa, a następnie analizowano uzyskane za ich pomocą wyniki. Zastosowano następujące oznaczenia:

- X oraz Y to zbiory podsumowań recenzji, gdzie w każdym zbiorze znajduje się dokładnie jedno podsumowanie opinii o danym produkcie, wygenerowane na podstawie innych recenzji, niż podsumowanie opinii o tym produkcie znajdujące się w drugim zbiorze,
- $d(s_{i,X}, s_{j,Y})$ to odległość pomiędzy podsumowaniem recenzji i -tego produktu pochodzącym ze zbioru X a podsumowaniem recenzji produktu j -tego pochodzącym z zbioru Y ,
- A to zbiór wszystkich aspektów, na temat których recenzenci się wypowiadali oceniając dany produkt lub usługę,
- $p_{i,a,o}$ to liczba pozytywnych, a $n_{i,a,o}$ negatywnych ocen, jakie zostały nadane aspektowi a , należącemu do produktu i w recenzjach należących do zbioru o ($o \in \{X, Y\}$),
- $r_{i,j,X,Y}$ to maksymalna liczba opinii nadanych dowolnemu aspektowi spośród podsumowań $s_{i,X}$ oraz $s_{j,Y}$.

Jak wspomniano, przygotowano kilka miar, za pomocą których przeprowadzono eksperymenty opisane w poprzedniej sekcji. Ostatecznie zdecydowano się na miarę zaprezentowaną w formule 5.4.

$$d_{(s_{i,X}, s_{j,Y})} = \sum_{a \in A} \frac{|(p_{i,a,X} - p_{j,a,Y})| + |(n_{i,a,X} - n_{j,a,Y})|}{2 * r_{i,j,X,Y}} \quad (5.4)$$

Miara ta została wypracowana przez autora pracy i wybrana ze względu na fakt, że pozwala na otrzymanie znormalizowanego wyniku niezależne od liczby recenzji nadanej danemu produktowi czy usłudze. Dla każdego aspektu, oblicza się różnice pomiędzy liczbą recenzji, w których dany aspekt został uznany jako zaleta w obu zbiorach (X i Y) oraz jako wada. Wartości bezwzględne otrzymanych różnic podlegają sumowaniu, którego rezultat dzieli się przez maksymalną liczbę opinii nadanych aspektowi w analizowanych podsumowaniach, co poddaje otrzymaną wartość normalizacji. Następnie, sumuje się wyniki opisanego działania dla wszystkich analizowanych aspektów.

Dla zobrazowania działania omówionej formuły, warto zamieścić przykład jej działania na danych zawartych w tabeli 5.3. Składnik funkcji podobieństwa dla aspektu „wytrzymałość” zostałby obliczony w następujący sposób:

$$\frac{|(p_{i,a,X} - p_{j,a,Y})| + |(n_{i,a,X} - n_{j,a,Y})|}{2r} = \frac{|(12 - 22)| + |(71 - 56)|}{2 * 102} = \frac{25}{204} = 0,1225 \quad (5.5)$$

dla:

- i = Samsung U700,
- a = wytrzymałość.

Współczynnik $r_{i,j,X,Y}$ ma wartość 102, ponieważ maksymalna liczba opinii nadanych dowolnemu aspektowi wśród analizowanych podsumowań to $99 + 3$ dla aspektu *waga* w Podsumowaniu 1. Aby uzyskać ostateczną odległość pomiędzy podsumowaniami, analogiczną operację należałoby następnie przeprowadzić dla wszystkich aspektów, a uzyskane wartości zsumować.

5.4.3 Procedura eksperymentu i uzyskane wyniki

Opisywany eksperyment nie wymagał przeprowadzania ręcznej adnotacji, w związku z czym przeprowadzany był na całym dostępnym korpusie recenzji (nie przeprowadzano losowania). Na początku wygenerowano tablice odległości (o strukturze analogicznej do tabeli 5.4) z wykorzystaniem przyjętej miary odległości i na posiadanym zestawie danych, składającym się z podsumowań 1529 produktów (w sumie 3058 podsumowań w dwóch zbiorach X i Y). Taką liczbę produktów uzyskano po odrzuceniu produktów posiadających tylko jedną recenzję (gdyż wtedy nie było możliwe podzielenie zbioru recenzji danego produktu na dwie części zawierające recenzje). Eksperyment przeprowadzono z wykorzystaniem skryptu przygotowanego w języku programowania Python.

Test istotności dla dwóch średnich

Pierwsze z badań polegało na przeprowadzeniu statystycznego testu istotności dla dwóch średnich, odpowiadających dwóm populacjom, gdzie na populację pierwszą składały się wartości odległości $d(s_{iX}, s_{jY})$ dla $i = j$ (ich średnią będziemy oznaczać jako μ_1), zaś na populację drugą wartości $d(s_{iX}, s_{jY})$ dla $i \neq j$ (średnią oznaczamy jako μ_2). Przyjęto następujące hipotezy:

- $H_0 : \mu_1 = \mu_2$ (średnie dla obu populacji są sobie równe),
- $H_1 : \mu_1 < \mu_2$ (średnia dla $i = j$ jest niższa niż dla pozostałych przypadków).

Ze względu na dużą liczbę obserwacji (odległości dla $i = j$ było 1529, zaś pozostałych odległości dla różnych wartości i i j 2336312, tj. są to wszystkie wartości z macierzy o rozmiarze 1529 na 1529 bez wartości na przekątnej tej macierzy), uzasadnione było wykorzystanie odchyłeń standardowych z prób jako szacunkowych wartości odchyłeń standardowych populacji oraz statystyki U do jej weryfikacji¹:

$$U = \frac{\overline{d_{i=j}} - \overline{d_{i \neq j}}}{\sqrt{\frac{s_{i=j}^2}{n_{i=j}} + \frac{s_{i \neq j}^2}{n_{i \neq j}}}} \quad (5.6)$$

gdzie:

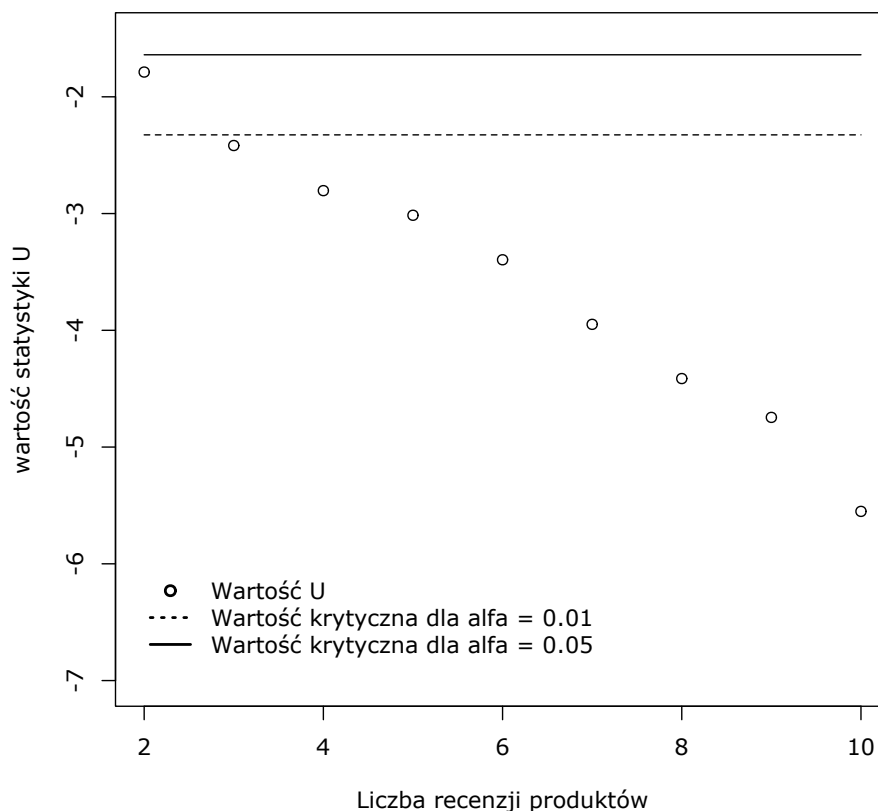
- $\overline{d_{i=j}}$ to średnia odległości, gdy $i = j$, zaś $\overline{d_{i \neq j}}$ to średnia odległości, gdy $i \neq j$,
- $s_{i=j}$ i $s_{i \neq j}$ to odchylenia standardowe wartości d dla $i = j$ i $i \neq j$,
- $n_{i=j}$ i $n_{i \neq j}$ to liczebności odpowiednich zbiorów.

Jako poziom istotności w teście przyjęto wartość $\alpha = 0.01$. W związku z tak postawionymi hipotezami oraz przyjętym poziomem istotności, jako wartość krytyczną przyjęto -2,326 (kwantyl rzędu 0,01 rozkładu normalnego $N(0, 1)$), zaś obszarem odrzucenia jest przedział liczbowy $(-\infty; -2, 326]$.

Dla miary przedstawionej w formule 5.4, średnią odległością pomiędzy podsumowaniami tych samych produktów (dla $i = j$) było 2,186 (przy odchyleniu standardowym 1,613), natomiast pomiędzy podsumowaniami różnych produktów było to 3,297 (odchylenie standardowe 1,04). W przeprowadzonym teście uzyskano wartość statystyki U na poziomie -26,673, co przy przyjętej wartości krytycznej (-2,326) spowodowało odrzucenie hipotezy zerowej na korzyść hipotezy, że

¹Ze względu na bardzo dużą liczbę obserwacji, pominięto również test na normalność analizowanych rozkładów.

średnia odległość pomiędzy podsumowaniami tych samych produktów (wygenerowanych na podstawie różnych recenzji) jest mniejsza niż pomiędzy podsumowaniami produktów różnych.



Rysunek 5.3: Wartości statystyki U w testach istotności dla dwóch średnich przy różnych populacjach produktów o niewielkiej liczbie recenzji. Źródło: opracowanie własne

Dodatkowo, przeprowadzono analogiczne badania dla podsumowań wygenerowanych na podstawie niewielkiej liczby recenzji. Miały one wskazać, czy w przypadku takich podsumowań występować będzie podobieństwo pomiędzy podsumowaniami tych samych produktów. Istniało tutaj bowiem ryzyko, że w przypadku niewielkiej liczby recenzji, subiektywne opinie różnych recenzentów mogą mieć zbyt duży wpływ na otrzymane podsumowania. Mogłoby to spowodować, że podsumowania tych samych produktów nie będą do siebie podobne.

Aby przeanalizować opisane zagadnienie, przeprowadzono testy dla produktów o liczbach recenzji od 2 do 10 (tj. przeprowadzono test dla produktów o liczbie recenzji 2, następnie dla produktów o liczbie recenzji 3 itd.). W każdym z przypadków, liczba takich podsumowań prze-

kraczała 30, co pozwalało na traktowanie odchyłeń standardowych z prób jako szacunkowych wartości odchyłeń standardowych populacji i korzystanie ze statystyki U . Warto zwrócić uwagę na fakt, że dla produktów o liczbie recenzji równej 2 oba podsumowania były generowane na podstawie tylko pojedynczych recenzji (czyli *de facto* były identyczne z tymi pojedynczymi recenzjami).

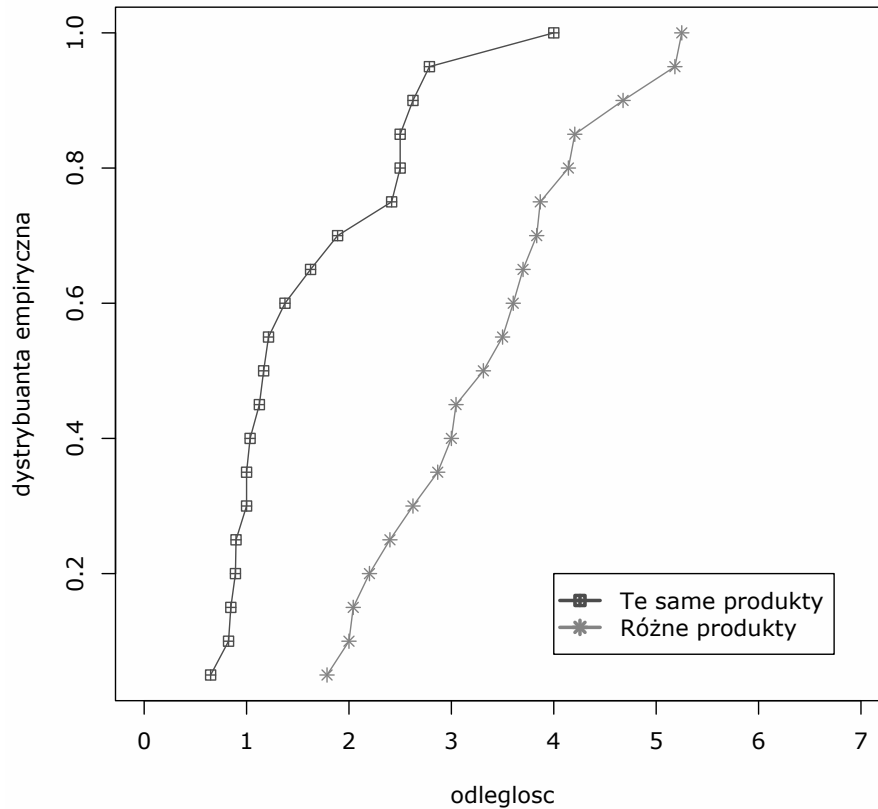
Otrzymane wyniki zaprezentowano na rysunku 5.3. Przedstawia on, jak zmieniała się wartość statystyki U dla produktów o różnej liczebności recenzji. Najważniejszymi wnioskami z przeprowadzonych testów są:

- dla produktów o liczbie recenzji co najmniej 3, za każdym razem hipoteza zerowa (o równości średnich w obu populacjach) została odrzucona,
- dla produktów o liczbie recenzji równej dwa (czyli gdy podsumowania były generowane na podstawie tylko jednej recenzji) dla $\alpha = 0,01$ nie było podstaw do odrzucenia H_0 , natomiast hipotezę tę można było odrzucić dla $\alpha = 0,05$,
- zaobserwowano wyraźny spadek wartości U wraz z coraz większą liczbą recenzji opisujących dany produkt.

Przeprowadzony eksperyment potwierdza postawioną hipotezę, że w przypadku niewielkiej liczby recenzji, podsumowania opinii o tym samym produkcie są coraz mniej do siebie podobne. Jednak wraz ze wzrostem liczby recenzji, coraz wyraźniej możemy zaobserwować, że produkty są postrzegane przez wielu recenzentów w spójny sposób.

Test Manna-Whitneya

Test Manna-Whitneya to nieparametryczny test istotności, który umożliwia przeprowadzanie porównań pomiędzy dwiema zbiorowościami pod względem pewnej cechy mierzalnej [Szwed 2009]. Wykorzystano ten test w celu ponownego potwierdzenia, że odległości pomiędzy podsumowaniami tych samych produktów są mniejsze, niż pomiędzy podsumowaniami produktów różnych. Na rysunku 5.4 porównano rozkłady dwóch próbek, z których jedna została wylosowana z populacji odległości pomiędzy podsumowaniami tych samych produktów, a druga z populacji odległości pomiędzy podsumowaniami produktów różnych. Analiza wykresu wykazuje, że rozkłady te znacząco się od siebie różnią (wykres skumulowanego rozkładu dla odległości pomiędzy różnymi podsumowaniami jest przesunięty w prawo). Test Manna-Whitneya ma za zadanie zweryfikować tę obserwację w sposób formalny.



Rysunek 5.4: Skumulowane rozkłady odległości pomiędzy podsumowaniami tych samych produktów oraz pomiędzy podsumowaniami produktów różnych. Źródło: opracowanie własne

W teście przyjęto następujące hipotezy:

- H_0 : rozkłady odległości pomiędzy podsumowaniami produktów z analizowanej populacji są takie same niezależnie od tego czy bierze się pod uwagę podsumowania tych samych, czy też różnych produktów,
- H_1 : w analizowanej populacji, odległości pomiędzy podsumowaniami tych samych produktów są statystycznie mniejsze, niż pomiędzy podsumowaniami produktów różnych.

Losowo wybrano dwie próbki, gdzie jedna z nich zawierała odległości pomiędzy podsumowaniami tych samych produktów, natomiast na drugą składały się odległości pomiędzy podsumowaniami produktów różnych. Obie próbki składały się z 20 obserwacji. Dla poziomu ufności $\alpha = 0,025$ wartością krytyczną testu jest 127, a obszar krytyczny obejmuje zakres poniżej tej

wartości.

Opisany test przeprowadzono czterokrotnie, za każdym razem ponownie przeprowadzając losowanie. Za każdym razem otrzymano wartości wpadające w obszar krytyczny, gdzie statystyki z testu miały wartości z zakresu od 75 do 88. Na podstawie uzyskanych statystyk, podjęto decyzję o odrzuceniu hipotezy zerowej na korzyść hipotezy, że odległości pomiędzy podsumowaniami tych samych produktów są statystycznie mniejsze, niż pomiędzy podsumowaniami produktów różnych.

5.4.4 Interpretacja uzyskanych wyników

Otrzymane wyniki wskazują na to, że faktycznie opinie wielu recenzentów o tym samym produkcie w analizowanym korpusie są ze sobą spójne. Wskazuje na to fakt, że podsumowania wielu opinii o danym produkcie, wygenerowane na podstawie rozłącznych kolekcji recenzji, są do siebie podobne. Odległości pomiędzy takimi podsumowaniami są statystycznie mniejsze, niż pomiędzy podsumowaniami różnych produktów. Można więc wysnuć wniosek, że na podstawie podsumowania wielu opinii o pewnym produkcie, możemy zyskać informację wskazującą na to, jaką opinię o tym samym produkcie prawdopodobnie mógł wygłosić inny recenzent. Oczywiście, pozyskana w ten sposób informacja nie jest pewna, jednak ponieważ wykazano spójność w opiniach wielu recenzentów, istnieje szansa, że w wielu przypadkach faktycznie takie podobieństwo opinii może wystąpić.

5.5 Podobieństwo podsumowań opinii wyrażanych w różnych formatach

W tej sekcji zaprezentowany zostanie eksperyment, mający na celu określenie, na ile podsumowania wszystkich opinii o pewnym produkcie wyrażonych za pomocą list zalet i wad są podobne do analogicznych podsumowań wygenerowanych na podstawie recenzji napisanych w języku naturalnym. Jak dowiedziono w sekcji 5.3, w ramach pojedynczej recenzji informacje zawarte w różnych sposobach wyrażania opinii w znacznym stopniu różnią się między sobą. W sekcji 5.4 dowiedziono natomiast, że podsumowania opinii różnych recenzentów o jednym produkcie są ze sobą spójne. W tej sekcji analizie poddane zostanie podobieństwo podsumowań wygenerowanych na podstawie różnych sposobów wyrażania opinii: recenzji tekstowych oraz list zalet i wad. Jeśli podsumowania recenzji w postaci list zalet i wad miałyby być wykorzystywane w ustalaniu polarności sformułowań służących do wyrażania opinii w recenzjach tekstowych, wynikiem opisanego

eksperymentu musi być wykrycie istotnego podobieństwa pomiędzy takimi podsumowaniami.

5.5.1 Podsumowania recenzji tekstowych a podsumowanie list zalet i wad

W tabeli 5.5 przedstawiono podsumowanie opinii recenzentów o wybranych aspektach telefonu Alcatel ot311, wyrażonych za pomocą różnych sposobów. Tabela ta ma analogiczną strukturę do tabeli 5.3, jednak tym razem zestawione są ze sobą podsumowania opinii wyrażonych na różne sposoby. Część tabeli pod nagłówkiem „Lista zalet i wad” prezentuje podsumowanie opinii o produkcie, wygenerowane na podstawie wszystkich recenzji w postaci list zalet i wad nadanych danemu produktowi. Analizując tę część tabeli wyraźnie widać, że poszczególne aspekty produktu są różnie postrzegane przez recenzentów - niektóre wyraźnie negatywnie (na przykład liczba funkcji czy też rozmiar wyświetlacza), inne mniej jednoznacznie, gdzie część recenzentów oceniła dany aspekt negatywnie, a część pozytywnie (np. żywotność baterii czy też jakość dźwięku), jeszcze inne natomiast jednoznacznie pozytywnie (w tym przypadku aspektem takim jest cena).

Aspekt	Recenzje tekstowe		Lista zalet i wad	
	Zalet	Wad	Zalet	Wad
obsługa	5	4	1	0
liczba funkcji	1	12	0	16
rozmiar wyświetlacza	0	8	0	18
rozdzielczość wyświetlacza	0	0	0	5
żywotność baterii	5	6	3	11
jakość dźwięku	0	1	1	6
cena	3	0	12	0

Tabela 5.5: Porównanie liczby recenzji, w których wymienione zostały wybrane aspekty produktu bądź w opisie tekstowym, bądź w liście zalet i wad. Porównanie przeprowadzono dla telefonu komórkowego Alcatel ot311 na podstawie strony <http://cokupic.pl/produkt/Alcatel-alcatel-ot311>. Dostęp: 1 lutego 2013

Z takim podsumowaniem zestawiono analogicznie przygotowane podsumowanie dla recenzji tekstowych (tabela 5.5, pod nagłówkiem „Recenzje tekstowe”). Podsumowanie takie przygotowano na podstawie adnotacji nadanych recenzjom tekstowym w sposób opisany w sekcji 5.2.2. Dla każdej recenzji przygotowano listę takich adnotacji, które traktowano tak samo jak zalety i

wady nadane przez recenzenta i które podsumowano w analogiczny sposób. W zaprezentowanym w tabeli przykładzie, w podsumowaniu recenzji tekstowych znajdują się inne wartości niż pod nagłówkiem „Lista zalet i wad”, jednak opinie wyrażane o poszczególnych aspektach za pomocą obu sposobów można uznać za podobne do siebie. Zarówno „liczba funkcji”, jak i „rozmiar wyświetlacza” zostały bardzo negatywnie ocenione (przy czym przy liczbie funkcji w recenzjach tekstowych pojawiła się jedna ocena pozytywna), zaś „żywołność baterii” ponownie ma ocenę mieszaną z przewagą ocen negatywnych (przy czym w przypadku recenzji tekstowych ta przewaga jest jednak wyraźnie mniejsza).

5.5.2 Koncepcja eksperymentów

Przeprowadzono formalną weryfikację podobieństwa pomiędzy podsumowaniami recenzji wyrażonych na oba sposoby. Eksperymenty w dużej mierze przeprowadzono w analogiczny sposób do eksperymentów opisanych w sekcji 5.4. Również tutaj przygotowano dwa zbiory podsumowań, gdzie tym razem na jeden zbiór składały się podsumowania wygenerowane na podstawie list zalet i wad, a na drugi na podstawie adnotacji nadanych recenzjom tekstowym. Następnie obliczono odległości pomiędzy wszystkimi elementami iloczynu kartezyjskiego obu zbiorów, otrzymując tablicę o analogicznej strukturze do tablicy przedstawionej w tabeli 5.4. W dalszej części sekcji stosowane będą analogiczne oznaczenia co w sekcji poprzedniej za wyjątkiem tego, że podsumowania będą oznaczane jako s_{iT} i s_{jL} , gdzie s_{iT} oznacza podsumowanie recenzji tekstowych i -tego produktu, zaś s_{jL} podsumowanie recenzji wyrażonych w postaci list zalet i wad produktu j -tego.

Eksperymenty przeprowadzono na zaadnotowanych recenzjach, jak opisano w sekcji 5.2.2. Były to następujące testy statystyczne:

- statystyczny test istotności, mający zweryfikować hipotezę o równości średnich odległości pomiędzy podsumowaniami recenzji tekstowych oraz recenzji w postaci zalet i wad dla tego samego produktu oraz średnich odległości pomiędzy podsumowaniami różnych produktów,
- test Manna-Whitneya, mający analogiczną funkcję, w celu potwierdzenia dokonanych obserwacji za pomocą testu nieparametrycznego.

5.5.3 Procedura eksperymentów i uzyskane wyniki

Badania zaprezentowane w tej sekcji przeprowadzono w analogiczny, jak opisane w sekcji 5.4.3. Wykorzystano również tę samą miarę podobieństwa pomiędzy podsumowaniami.

Test istotności dla dwóch średnich

Jako pierwsze badanie przeprowadzono statystyczny test istotności dla średnich odległości $d_2(s_{iT}, s_{jL})$ dla próbki, w której $i = j$ (czyli gdzie porównywane są podsumowania tego samego produktu) oraz drugiej, na którą składają się pozostałe przypadki. Obie analizowane próbki miały po 16 obserwacji (dla jednej z nich było dostępne dokładnie tyle obserwacji, zaś druga dostępną próbkę przycięto losowo do tego rozmiaru). Ze względu na niewielką liczbę obserwacji w próbkach, nie było możliwe traktowanie odchyłeń standardowych z prób jako szacunkowych wartości odchyłeń standardowych populacji. W celu ustalenia, jakiej statystyki należy użyć w teście, sprawdzono następujące cechy uzyskanego rozkładu odległości:

- za pomocą testu Kołmogorowa-Smirnowa zweryfikowano założenie normalności rozkładów w obu populacjach. Test nie pozwolił na odrzucenie hipotezy o takiej normalności w żadnej z populacji (uzyskano wartość statystyk na poziomie 0,167 i 0,135 przy wartości krytycznej wynoszącej 0,328 z prawostronnym obszarem odrzucenia dla poziomu ufności równego 0,05),
- przeprowadzono test F (Fishera) na istotność dla równości wariancji z obu populacji, dla poziomu ufności na poziomie 0,05. Wartość statystyki była na poziomie 1,125 przy wartości krytycznej 2,4 (dla 15 i 15 stopni swobody) i prawostronnym obszarze odrzucenia. W związku z tym, nie było podstaw, aby odrzucić hipotezę o równości wariancji w obu populacjach.

Upewniwszy się co do założeń testu istotności dwóch średnich, przystąpiono do przeprowadzenia właściwego testu, mającego za zadanie ustalić, czy średnia odległość pomiędzy podsumowaniami opinii wyrażonych na różne sposoby dla tych samych produktów jest niższa niż odległość pomiędzy podsumowaniami opinii o różnych produktach.

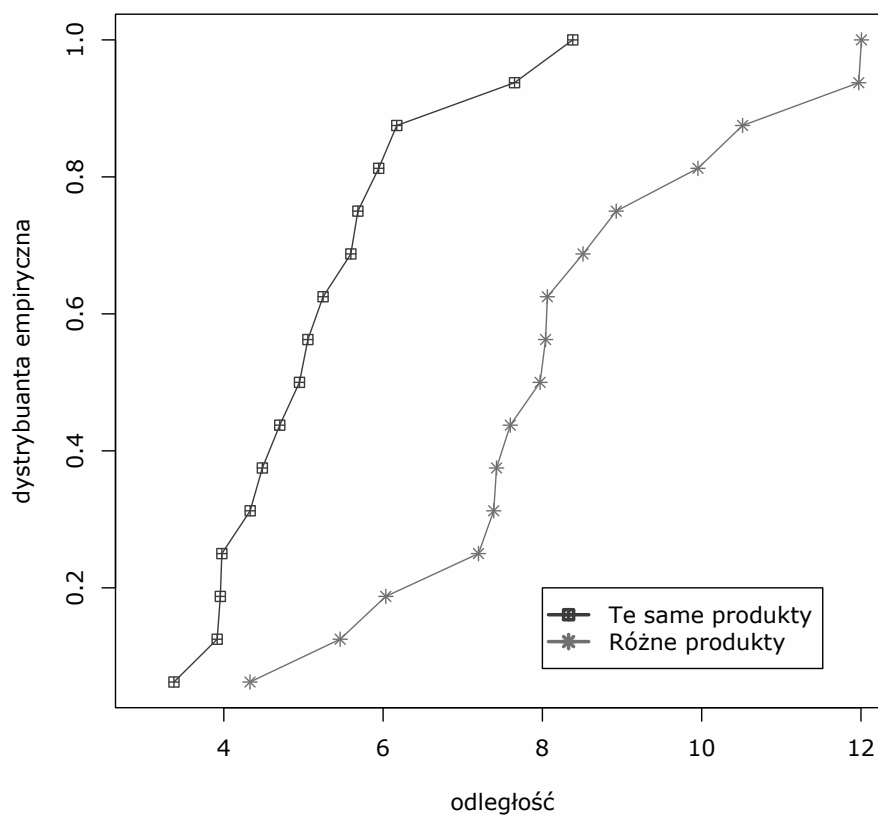
Średnią podobieństw pomiędzy podsumowaniami tych samych produktów oznaczamy jako μ_1 , zaś pomiędzy podsumowaniami produktów różnych jako μ_2 . Przyjęto następujące hipotezy:

- $H_0 : \mu_1 = \mu_2$ (średnie dla obu populacji są sobie równe),
- $H_1 : \mu_1 < \mu_2$ (średnia dla $i = j$ jest niższa niż dla pozostałych przypadków).

Ze względu na niewielką liczbę obserwacji oraz normalność rozkładów i równość wariancji w obu populacjach, hipotezę weryfikowano za pomocą statystyki t-Studenta. Otrzymaną wartością statystyki było -6,514 dla wartości krytycznej przy poziomie ufności $\alpha = 0,01$ i 30 stopniach swobody równej -2,457. W związku z tym, hipotezę o równości średnich w obu populacjach należy

odrzuć. Oznacza to, że odległości pomiędzy podsumowaniami tych samych produktów, gdzie jedno podsumowanie było wygenerowane na podstawie list zalet i wad, a drugie na podstawie adnotacji wypowiedzi tekstowych, są statystycznie niższe niż pomiędzy analogicznymi podsumowaniami produktów różnych.

Test Manna Whitneya



Rysunek 5.5: Skumulowane rozkłady odległości pomiędzy podsumowaniami tych samych produktów oraz pomiędzy podsumowaniami produktów różnych, gdzie podsumowania zostały wygenerowane na podstawie różnych sposobów wyrażania opinii – wypowiedzi tekstowej oraz listy zalet i wad. Źródło: opracowanie własne

Na rysunku 5.5 zaprezentowano rozkłady odległości pomiędzy podsumowaniami recenzji tekstowych oraz recenzji w postaci zalet i wad. Na jedną próbkę składały się odległości pomiędzy podsumowaniami tych samych produktów, a na drugą odległości pomiędzy podsumowaniami pro-

duktów różnych. Dla zaprezentowanych danych ponownie przeprowadzono test Manna Whitneya.

W teście przyjęto następujące hipotezy:

- H_0 : w analizowanej populacji rozkłady odległości pomiędzy podsumowaniami produktów, wygenerowanymi na podstawie recenzji tekstowych oraz list zalet i wad, są takie same niezależnie od tego, czy bierze się pod uwagę podsumowania tych samych, czy różnych produktów,
- H_1 : w analizowanej populacji odległości pomiędzy podsumowaniami tych samych produktów (gdzie jedno podsumowanie jest wygenerowane na podstawie recenzji tekstowych, a drugie na podstawie list zalet i wad) są stochastycznie mniejsze, niż pomiędzy podsumowaniami produktów różnych.

Obie próbki miały rozmiar 16 podsumowań. Dla poziomu ufności $\alpha = 0,025$ wartością krytyczną testu jest 75, a obszar krytyczny obejmuje zakres poniżej tej wartości. Test przeprowadzono czterokrotnie, za każdym razem przeprowadzając osobne losowanie do próbki odległości pomiędzy podsumowaniami różnych produktów. Średnią wartością statystyki testowej było 25,25, a poszczególne wartości statystyki wahały się między 12 a 40. W związku z tym, odrzucono hipotezę zerową na rzecz hipotezy, że odległości pomiędzy podsumowaniami tych samych produktów (gdzie jedno podsumowanie jest wygenerowane na podstawie recenzji tekstowych, a drugie na podstawie list zalet i wad) są stochastycznie mniejsze niż pomiędzy podsumowaniami produktów różnych.

5.5.4 Interpretacja wyników

Wyniki eksperymentów w tej sekcji potwierdzają obserwację, że opinie wielu recenzentów o tym samym produkcie są ze sobą spójne. Wykazano, że w analizowanej próbie podsumowania opinii o tym samym produkcie, ale wyrażone na różne sposoby (gdzie jedno podsumowanie zostało wygenerowane na podstawie recenzji tekstowej, a drugie na podstawie list zalet i wad) są do siebie podobne. Potwierdza to obserwację dokonaną w sekcji 5.4 o spójności opinii recenzentów o produktach. Można więc oczekiwać, że na podstawie podsumowania wielu opinii o pewnym produkcie, wyrażonych w postaci list zalet i wad, można z pewnym prawdopodobieństwem przewidywać opinię, jaką wyrazi recenzent w wypowiedzi tekstowej, oceniającej ten sam produkt.

5.6 Podsumowanie

Po szczegółowym zaprezentowaniu koncepcji proponowanej metody w rozdziale 4, rozdział piąty miał na celu analizę danych przez tę metodę wykorzystywanych w celu wykazania, że prace w proponowanym kierunku są uzasadnione. Najważniejszymi wnioskami uzyskanymi z przeprowadzonych eksperymentów są:

- W ramach pojedynczej recenzji, informacje wyrażone w różnych częściach recenzji (wyrażonych na różny sposób) mogą się w znacznym stopniu różnić między sobą; przykładowo, w obu częściach recenzji oceniane są inne aspekty. Oznacza to, że mogą istnieć sytuacje, gdy nie jest możliwe wykorzystanie posiadanych danych do identyfikacji polarności na zasadach nadzorowania odległego. Jednocześnie, w analizowanym korpusie bardzo rzadko w ramach pojedynczych recenzji można natknąć się na sprzeczne informacje.
- Produkty są postrzegane przez recenzentów w spójny sposób. Dzięki temu, znając podsumowanie opinii wielu recenzentów o danym produkcie można z pewnym prawdopodobieństwem przewidzieć, jaką opinię może o tym produkcie wyrazić inny recenzent, a tym samym jaka jest polarność sformułowania użytego przez niego do wyrażenia opinii.
- Informacje zawarte w podsumowaniach recenzji wyrażonych w postaci list zalet i wad oraz tekstu w języku naturalnym są do siebie podobne. Oznacza to, że podsumowania recenzji wyrażonych w jednym formacie mogą być wykorzystane w charakterze wskazówek dotyczących opinii wyrażanych przez recenzentów w recenzjach wyrażonych w formacie innym.

Uzyskane rezultaty pozwalają oczekiwać, że zaproponowane w pracy podejście może w istocie pozwolić na poprawę skuteczności istniejących metod określania polarności sformułowań wykorzystywanych do wyrażania opinii o aspektach produktów i usług.

Rozdział 6

Ekstrakcja sformułowań służących do wyrażania opinii i wykorzystanie częściowo strukturyzowanych recenzji konsumenckich dla identyfikacji ich polarności

Celem rozdziału jest zaprezentowanie stworzonego rozwiązania implementującego pomysł opisany w rozdziale 4. Ten rozdział, wraz z rozdziałem 7, w którym opisano ewaluację wypracowanej metody, ma na celu wykazanie tezy postawionej we Wprowadzeniu do pracy i osiągnięcie jej drugiego i trzeciego celu szczegółowego.

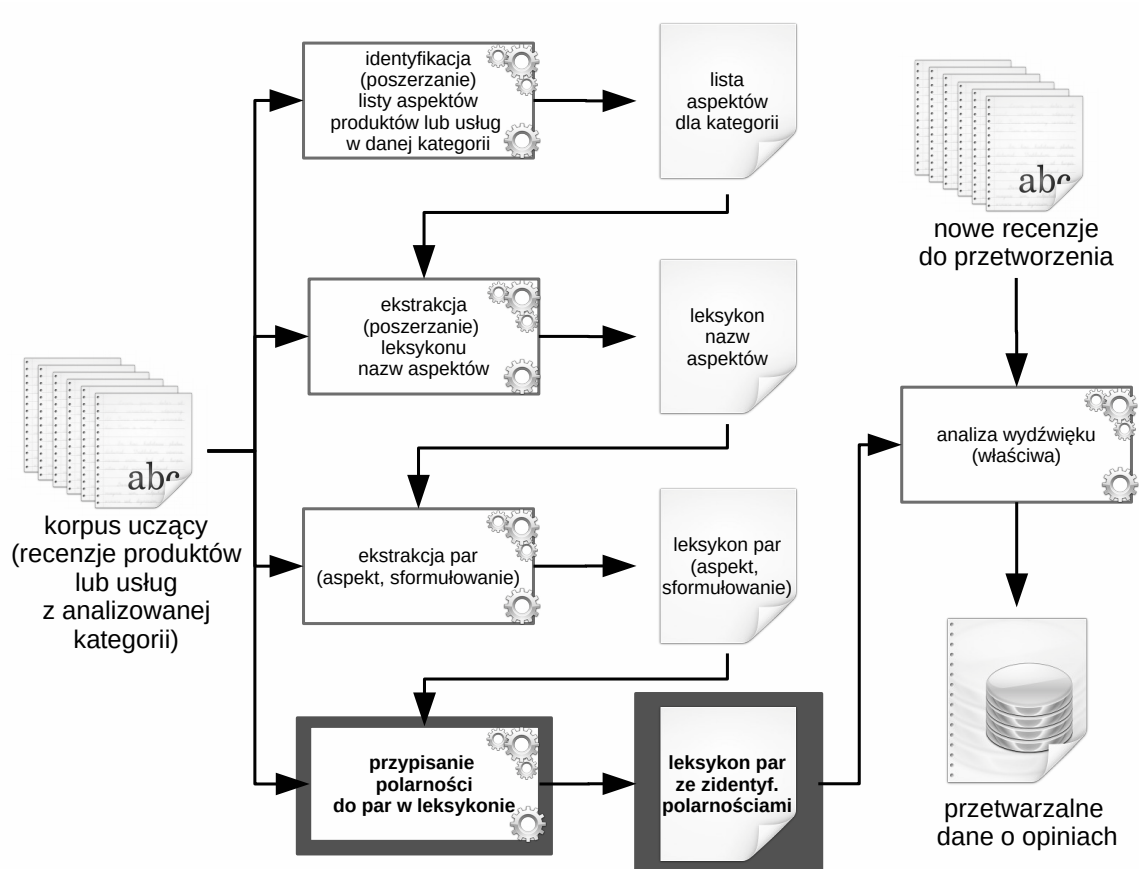
W kolejnych sekcjach opisano poszczególne etapy pracy i ich rezultaty, wraz z wyjaśnieniem, jaką rolę mają dla wykazania tezy pracy.

6.1 Opis rozwiązania

Głównym obszarem zainteresowania prezentowanej pracy jest identyfikacja polarności sformułowań wyrażających opinie dla różnych aspektów produktów lub usług, do których te sformułowania mogą się odnosić. Opisywane zadanie nie może jednak być rozpatrywane niezależnie od innych kroków w procesie analizy wydzwięku. Na rysunku 6.1 zaprezentowano schemat procesu analizy

wydźwięku wykorzystywany w pracy, w którym identyfikacja polarności jest jedynie jednym z etapów.

Zgodnie z rysunkiem 6.1, analiza wydźwięku składa się z następujących po sobie etapów (reprezentowanych jako prostokąty z ikonami kół zębatach), w wyniku których powstają pewne zasoby (reprezentowane w postaci kartki papieru). Każdy z etapów procesu, za wyjątkiem właściwej analizy wydźwięku, wykorzystuje korpus uczący, to jest kolekcję recenzji produktów lub usług z analizowanej kategorii. Przynajmniej część recenzji musi być wyrażona w formacie umożliwiającym wygenerowanie podsumowań wielu opinii o produktach i usługach. Dodatkowo, wejściem do każdego z kroków procesu (za wyjątkiem etapu pierwszego) jest również zasób będący rezultatem poprzedniego kroku. Przykładowo, krok *ekstrakcja par (aspekt, sformułowanie)*, oprócz korpusu uczącego, wymaga do działania leksykonu nazw aspektów.



Rysunek 6.1: Schemat opracowanego rozwiązania. Źródło: opracowanie własne

Omawiany schemat można podzielić na dwie główne części:

- naukę – składają się na nią cztery procesy przedstawione po lewej stronie rysunku, z których

każdy wykorzystuje korpus uczący; finalnym rezultatem jest leksykon wydźwięku,

- wykorzystanie – proces analizy wydźwięku, gdzie przetwarzane są nowe recenzje (niewykorzystywane podczas nauki); analiza wydźwięku działa z wykorzystaniem leksykonu par z przypisanymi polarnościami wygenerowanego podczas nauki.

Główny obszar zainteresowania prezentowanej pracy, zgodnie z pomysłem opisanym w rozdziale 4 i tezą sformułowaną we Wprowadzeniu, na diagramie można umiejscowić w operacji „przypisanie polarności do par w leksykonie”. W związku z tym, wkładem naukowym tej pracy będzie usprawnienie tego właśnie procesu. Należy zwrócić uwagę na fakt, że procesy, które na schemacie 6.1 mają miejsce przed przypisywaniem polarności, stanowią odrębne problemy badawcze.

W kolejnych sekcjach rozdziału opisano kolejno sposób realizacji poszczególnych etapów z fazy nauki z diagramu 6.1. Jako przykład korpusu w trakcie opisywania wypracowanego rozwiązania traktowane będą recenzje pobrane z portalu `cokupic.pl`, opisane wcześniej w sekcji 5.2.2. Drugim korpusem, na którym opisana metoda była testowana, były recenzje restauracji pobrane z portalu `gastronauci.pl` (patrz sekcja 4.3.1). Za pomocą tych danych, opracowane rozwiązanie będzie podlegało ewaluacji w kolejnym rozdziale.

Dla ułatwienia lektury rozdziału, w większości przykładów zamieszczonych w tekście będziemy się odwoływać do korpusu recenzji telefonów komórkowych.

6.2 Wymagania dla metody

Ze względu na bogactwo języka naturalnego i stopień skomplikowania wielu zadań związanych z jego przetwarzaniem, konieczne jest dokładne zdefiniowanie planowanego zakresu prac i zadań, które opracowywana metoda ma rozwiązywać. Jak wspomniano w sekcji 3.1, w istniejących pracach z zakresu analizy wydźwięku skupiano się na wybranych elementach piątki reprezentującej opinię (np. identyfikacja jedynie wydźwięku przypisanego do całego dokumentu, bez analizy opinii o bytach i aspektach w nim wspomnianych).

W analizie wydźwięku produkty, usługi i ich aspekty mogą być oceniane z wykorzystaniem skomplikowanych sformułowań, których prawidłowe przetworzenie wymagałoby wyposażenia systemu w szeroką wiedzę o świecie i skomplikowane mechanizmy wnioskowania. Przygotowanie uniwersalnego rozwiązania, które byłoby w stanie przypisać właściwy wydźwięk do dowolnego tekstu, pozostaje aktualnie nieosiągalne.

Z tego powodu, szczegółowo zdefiniowano wymagania dla przygotowywanego rozwiązania i określono, jakie zagadnienia znajdują się poza obszarem zainteresowania rozprawy. Wymagania te dotyczą procesów wyszczególnionych na rysunku 6.1 na stronie 114. W trakcie ich definiowania, wzorowano się na istniejących pracach, które miały cel podobny do prezentowanej pracy (zostały one opisane w rozdziale 3). W związku z tym, sformułowano następujące główne wymagania dotyczące przygotowywanej metody:

1. wynikiem prac ma być metoda identyfikacji polarności dla sformułowań z zadanego leksykonu, działająca na podstawie analizy podsumowań wielu opinii o dobrach i ich aspektach, opisywanych w recenzjach, w których te sformułowania występowały,
2. identyfikacja polarności ma być przeprowadzana na poziomie poszczególnych aspektów recenzowanych produktów i usług, tj. faktycznie polarność przypisywana jest do pary (*sformułowanie, aspekt*),
3. leksykon sformułowań służących do wyrażania opinii ma być generowany automatycznie na podstawie analizy korpusu,
4. otrzymany leksykon, z polarnościami przypisanymi przez zaproponowaną metodę, ma dawać lepsze rezultaty analizy wydźwięku, niż leksykon z polarnościami przypisanymi z wykorzystaniem istniejących metod identyfikacji polarności z nurtu nadzorowania odległego.

Jednocześnie, sformułowano następujące ograniczenia w obszarze zainteresowania prezentowanej pracy:

1. konstrukcja leksykonu aspektów i ich nazw dla recenzowanych dóbr będzie wykonywana manualnie (tj. nie zostaną wykorzystane automatyczne metody realizowania tego zadania),
2. leksykon sformułowań służących do wyrażania opinii ma składać się z określonych, zdefiniowanych typów sformułowań, zgodnych ze zdefiniowanym wzorcem (np. sformułowania zawierające wyrazy o określonych częściach mowy); zgodnie z wymaganiem trzecim zdefiniowanym powyżej, sformułowania zgodne ze zdefiniowanym typem mają być ekstrahowane i dodawane do leksykonu w sposób automatyczny,
3. analiza wydźwięku na poziomie aspektów dotyczy wyłącznie sytuacji, gdy mają miejsce odwołania do aspektów wprost, tj. gdy pojawia się nazwa danego aspektu oraz sformułowanie z leksykonu sformułowań służących do wyrażania opinii,

4. w trakcie przetwarzania recenzji przyjęte jest założenie, że są to recenzje pojedynczego produktu (nie są to np. recenzje porównawcze), a recenzenci nie korzystają z wypowiedzi ironicznych.

Należy zaznaczyć, że wprowadzone ograniczenia nie wpływają na istotność rezultatów prezentowanej rozprawy. Wypracowane metody działałyby również w sytuacji, gdyby powyższe ograniczenia zostały zniesione; jedynymi powodami wprowadzenia wymienionych ograniczeń było zapewnienie odpowiedniej zwięzłości pracy oraz zapewnienie możliwości poświęcenia odpowiedniej uwagi zagadnieniom koniecznym do wykazania tezy prezentowanej rozprawy.

6.3 Przygotowanie leksykonu aspektów i ich nazw

6.3.1 Identyfikacja aspektów

Zgodnie z ideą nadzorowania odległego, aby możliwe było przeprowadzenie nauki polarności sformułowań na poziomie poszczególnych aspektów recenzowanych dóbr, konieczne jest posiadanie przetwarzalnych informacji o opiniach recenzentów o tych dobrach i aspektach. Przykładowo, opinie o konkretnym aspekcie powinny być wyrażane w postaci skali punktowej. W związku z tym, lista aspektów, dla których może być przeprowadzona analiza, jest ograniczona do aspektów, które mogą być oceniane w danym korpusie za pomocą formatu umożliwiającego łatwe ich przetworzenie. W przypadku korpusu recenzji telefonów komórkowych z portalu `cokupic.pl` będą to te aspekty, które mogą być oceniane za pomocą list zalet i wad, tak jak to zostało opisane w sekcji 5.2.2. Analogicznie, w przypadku recenzji restauracji na portalu `gastronauci.pl` taka lista również jest dostępna (są to aspekty, które można oceniać za pomocą skali punktowej).

Oczywiście, nie w każdym przypadku dysponujemy gotową, ograniczoną listą aspektów. W przypadku opisanym w scenariuszu w sekcji 4.3.3, lista aspektów nie była z góry ograniczana, gdyż możliwe jest tu wygenerowanie podsumowania opinii o dowolnym aspekcie na podstawie analizy wielu recenzji tekstowych danego produktu. Aspekty mogłyby tu być pozyskiwane automatycznie z tekstu. Istnieje wiele prac, których celem była właśnie identyfikacja aspektów produktów ocenianych w wypowiedziach tekstowych, co zostało opisane w sekcji 3.2, czyli przykładowo prace [Hu i Liu 2004a,b; Popescu i Etzioni 2005].

W związku z tym, nie przygotowano rozwiązania służącego do automatycznej identyfikacji aspektów ocenianych produktów i usług. W przykładowym scenariuszu przetwarzania recenzji z

portalu `cokupic.pl` zdefiniowano 22 aspekty, analogicznie jak to miało miejsce w sekcji 5.2.2. Lista zdefiniowanych aspektów znajduje się w załączniku B na stronie 179.

6.3.2 Leksykon nazw aspektów

Jak wspomniano w sekcji 6.2, w przygotowywanym rozwiązaniu analizowane są wyłącznie odniesienia wprost do aspektów, to jest takie, w których w tekście pojawiło się słowo będące nazwą danego aspektu. Jako nazwę aspektu rozumiemy rzeczownik lub frazę rzeczownikową używaną przez recenzentów w celu odniesienia się w tekście do danego aspektu. Jak opisano w sekcji 3.2.1, każdy aspekt może mieć wiele takich nazw. W związku z tym konieczne jest dla każdego aspektu przygotowanie leksykonu jego możliwych nazw.

Istnieją prace, które dla zadanej taksonomii aspektów mają na celu ekstrakcję ich nazw, np. [Carenini, Ng i Zwart 2005; Liu, Hu i Cheng 2005]. Podobnie jednak jak w przypadku identyfikacji aspektów, zadanie to nie leży w głównym obszarze zainteresowań pracy. W związku z tym, podjęto decyzję o ręcznym przygotowaniu leksykonów nazw aspektów. Procedurę tę przeprowadzono na drodze analizy przez dwóch adnotatorów korpusu 450 losowych recenzji pochodzących z trzech portali z recenzjami telefonów komórkowych (`cokupic.pl`, `opineo.pl` i `bozo.pl`). Zadaniem adnotatorów było przeczytanie recenzji i w każdej z nich zaznaczenie wyrazów, które ich zdaniem są nazwami aspektów produktów pochodzących z podanej listy. W tym celu zdefiniowano wskazówki, którymi mieli się kierować adnotatorzy (patrz załącznik B).

Obaj adnotatorzy (nazwijmy ich adnotatorem A oraz adnotatorem B), niezależnie od siebie, analizowali recenzje i oznaczali nazwy aspektów. Wyrazy oznaczone przez adnotatorów zostały następnie automatycznie pobrane z korpusu i pogrupowane według aspektu, którego dotyczyły. Jako prawidłowe adnotacje traktowano wyłącznie te, które zostały nadane identycznie przez obu adnotatorów.

W sumie adnotatorzy dokonali dokładnie 1000 adnotacji, przy czym:

- 513 adnotacji zostało nadanych identycznie przez obu adnotatorów,
- 187 adnotacji zostało nadanych przez adnotatora A i pominiętych przez adnotatora B,
- 227 adnotacji zostało nadanych przez B i pominiętych przez A,
- w 73 przypadkach adnotacje A i B nachodziły na siebie (miały inny zakres, przykładowo adnotator A oznaczył dodatkowe słowo w stosunku do adnotatora B).

Wiele z adnotacji oznacza te same wyrazy użyte w różnych miejscach. Tym samym, lista unikalnych nazw aspektów jest znacznie krótsza i wynosiła 109 pozycji. Zostały one podzielone na osobne leksykony dla każdego aspektu. Następnie przeanalizowano otrzymane leksykony i wzbogacono je o dodatkowe pozycje, takie jak wszystkie odmienne formy znajdujących się w nich nazw. Co ważne, dla części aspektów adnotatorzy nie oznaczyli żadnych nazw (były to *awaryjność*, *gps*, *kompatybilność* i *serwis*). Aspekty te wykluczono z dalszego przetwarzania.

W ten sposób otrzymano leksykony nazw aspektów. Dzięki temu, możliwe było przygotowanie implementacji dalszych zadań w procesie nauki.

6.4 Ekstrakcja sformułowań służących do wyrażania opinii

Posiadając listę aspektów oraz leksykony ich nazw, możliwe było przystąpienie do prac nad ekstrakcją sformułowań wyrażających opinie. Celem tego procesu była identyfikacja sformułowań, za pomocą których recenzenci wyrażali w korpusie opinii o poszczególnych aspektach recenzowanych produktów i usług. Wyjściem z tego zadania byłby leksykon, który dla każdego aspektu definiowałby sformułowania, za pomocą których dany aspekt może być oceniany. Co ważne, na tym etapie w leksykonie nie znajdują się jeszcze informacje o tym, czy dane sformułowanie jest wykorzystywane do wyrażania pozytywnej, negatywnej czy neutralnej opinii o recenzowanym aspekcie.

W istniejących pracach prezentowano różne podejścia do zagadnienia ekstrakcji sformułowań wyrażających opinie. Wykorzystywano tu m.in.:

- ręcznie skonstruowane reguły (wzorce) ekstrakcji, które identyfikowały określone sekwencje słów ([Grefenstette i in. 2006; Wawer 2012]) albo słów i części mowy ([Popescu i Etzioni 2005; Rill i in. 2012; Turney 2002]),
- wyniki głębokiego parsowania języka, gdzie zdania były rozpisywane w postaci drzewa i gdzie wybierano określone poddrzewa [Lu i in. 2011; Qiu i in. 2009].

Rozwiązanie omawianego zadania było konieczne dla potrzeb dalszych eksperymentów, jednak nie leży ono w głównym obszarze zainteresowań pracy. Dla potrzeb realizacji tego zadania przygotowano i przeprowadzono ewaluację autorskiego rozwiązania dot. ekstrakcji sformułowań do leksykonu.

6.4.1 Analiza korpusu

Pierwszym krokiem w pracach nad ekstrakcją sformułowań służących do wyrażania opinii była analiza korpusu recenzji konsumenckich i tego, w jaki sposób recenzenci wyrażali swoje opinie o dobrach i ich aspektach. Ponownie, dwóch adnotatorów analizowało 450 dokumentów tekstowych, będących recenzjami telefonów komórkowych, pobranych z trzech portali z recenzjami. Adnotacja przebiegała zgodnie ze sformułowanymi wskazówkami (patrz załącznik C na stronie 185). Celem procesu adnotacji było zebranie bazy przykładowych sformułowań wykorzystywanych przez recenzentów w celu wyrażenia opinii o którymś ze zdefiniowanych aspektów recenzowanych produktów (w tym przypadku telefonów komórkowych). Adnotowano fragmenty tekstów, w których znajdowała się nazwa ocenianego aspektu oraz słowo (lub dłuższe sformułowanie) wyrażające opinię o danym aspekcie. Każda adnotacja zawierała następujące informacje:

- jaki aspekt jest oceniany,
- jaka jest polarność danej frazy – pozytywna czy negatywna (przy czym na tym etapie ta informacja nie była wykorzystywana).

Dodatkowo, adnotatorzy mieli za zadanie oznaczanie wszystkich kontekstów leksykalnych, które w analizowanych recenzjach zmieniały znaczenie fraz, za pomocą których recenzenci wyrażali opinie. Prosty przykładem może być tutaj negacja, np. „nie powiem, żeby bateria trzymała długo”.

Od obu adnotatorów otrzymano w sumie 855 adnotacji, które poddano analizie w celu wypracowania metody ekstrakcji sformułowań wyrażających opinie. W trakcie analizy przeprowadzono następujące czynności:

- zidentyfikowano często występujące sekwencje wyrazów o określonych częściach mowy i pozycji z leksykonu, za pomocą których recenzenci wyrażają opinie o aspektach recenzowanych produktów,
- zidentyfikowano często występujące szersze konteksty leksykalne, w których takie frazy występują,
- określono, jakie wyrazy często powodują zmianę wydźwięku fraz w tekście.

6.4.2 Przygotowane rozwiązanie

Z przeprowadzonej analizy korpusu wyciągnięto dwa główne wnioski:

- wypowiedzi w korpusie cechują się znaczną liczbą literówek, błędów ortograficznych i błędów gramatycznych. Z tego względu, odrzucono możliwość wykorzystania głębokiego parsowania języka w celu identyfikacji konstrukcji gramatycznych określonego typu,
- frazy, w których recenzenci wyrażają swoje opinie, często zamieszczone są w podobnych do siebie kontekstach. Konteksty te składają się z określonych słów i części mowy (np. wielu przymiotników i fraz przymiotnik – rzeczownik), a zdefiniowanie wszystkich dokładnych sekwencji, w jakich występują słowa i części mowy w otoczeniu fraz służących do wyrażania opinii, byłoby bardzo czasochłonne.

Z uwagi na drugi z opisanych powyżej wniosków, podjęto decyzję o wykorzystaniu podejścia regułowego do ekstrakcji fraz potencjalnie używanych przez recenzentów do wyrażania opinii, w trakcie którego nie analizowano szerszego kontekstu, w którym frazy te występują, natomiast szerszy kontekst jest analizowany z wykorzystaniem klasyfikatora wygenerowanego na drodze uczenia maszynowego. Przygotowane reguły ekstrakcji miały być ogólne, a tym samym miały identyfikować wiele potencjalnych fraz, w których recenzenci mogli wyrażać opinie. Z tego powodu istniało ryzyko błędnej identyfikacji dużej liczby fraz, które zawierały sekwencję części mowy i słów zgodną ze zdefiniowanymi regułami, jednak faktycznie nie wyrażano za ich pomocą opinii (na co wskazywać może kontekst, w jakim fraza wystąpiła). Dodatkowy etap filtrowania miał odrzucać takie właśnie niewłaściwie zidentyfikowane frazy. Jak zaznaczono, w szerszych kontekstach fraz, w których recenzenci wyrażali swoje opinie, często występowały określone części mowy i słowa, jednak nie dało się ustalić ogólnych sekwencji, w których zazwyczaj występują. Stąd, podjęto decyzję o potraktowaniu kontekstu jako zbioru (pozbawionej porządku kolekcji) słów i fraz, na podstawie którego generowano określone cechy wykorzystywane w uczeniu maszynowym i później w klasyfikacji, która każdą frazę, na podstawie jej kontekstu, przypisywała do jednej z dwóch klas: fraza zawierająca sformułowanie wyrażające opinię (sformułowanie było dodawane do leksykonu) i fraza niezawierająca takiego sformułowania (odrzucała z dalszego przetwarzania).

W związku z tym, przygotowywane rozwiązanie zostało podzielone na trzy główne fazy:

1. przygotowawczą,
2. regułową,
3. statystyczną (wykorzystującą uczenie maszynowe).

W kolejnych podsekcjach szczegółowo omówiono powyższe fazy wypracowanego rozwiązania.

Faza przygotowawcza

Faza przygotowawcza miała na celu doprowadzenie dokumentów tekstowych do odpowiedniej postaci, która byłaby następnie wykorzystywana w kolejnych fazach ekstrakcji.

Najważniejszym krokiem w omawianej fazie było rozpoznanie części mowy wszystkich wyrazów występujących w tekstach i przypisanie na tej podstawie adnotacji do poszczególnych wyrazów. W tym celu wykorzystano program TaKIPI¹, będący tagerem języka polskiego. Zgodnie z dokumentacją, program działa ze skutecznością 93,4% [Piasecki 2007].

Program TaKIPI zapisywał przetworzone dokumenty w plikach XML, w których znajdowały się informacje o częściach mowy poszczególnych wyrazów oraz o ich formach podstawowych.

Faza regułowa

W fazie regułowej, głównym celem była identyfikacja fragmentów tekstów, za pomocą których recenzenci prawdopodobnie wyrażają opinie o aspektach produktów i usług. Aby było to możliwe, konieczne było stworzenie odpowiednich reguł ekstrakcji, to jest zidentyfikowanie typowych kontekstów leksykalnych, za pomocą których recenzenci wyrażają swoje opinie, a następnie zapisanie ich w wybranym formalizmie [Wieloch 2011].

Lp.	Sekwencja	Przykład
1	(ADJ PACT) ASPEKT	przystępna cena
2	(NUM ADV) (ADJ PACT) ASPEKT	długo trzymająca bateria
3	ADV (NUM ADV) (ADJ PACT) ASPEKT	bardzo wiele przydatnych funkcji
4	ASPEKT (<i>jest są</i>) (ADJ PACT)	sygnał jest silny
5	ADJ PREP ASPEKT	wygodny w obsłudze

Tabela 6.1: Zidentyfikowane sekwencje słów, nazw aspektów i części mowy o największym pokryciu dla analizowanego korpusu, dla których przygotowano reguły ekstrakcji. Źródło: opracowanie własne

W tabeli 6.1 zaprezentowano sekwencje, które w trakcie analizy korpusu zostały zidentyfikowane jako często wykorzystywane przez recenzentów do wyrażania opinii. Podjęto decyzję o

¹<http://nlp.pwr.wroc.pl/takipi>

ograniczeniu przeprowadzanych prac właśnie do tych reguł i pomijaniu innych sekwencji, za pomocą których recenzenci mogą wyrażać swoje opinie. Jest to uzasadnione z kilku powodów:

- proponowana metoda identyfikacji polarności, której wypracowanie jest głównym celem pracy, działa niezależnie od typu fraz w leksykonie, w związku z czym przeprowadzenie eksperymentów na frazach opisanych zgodnie ze zdefiniowanymi regułami jest wystarczające do wykazania tezy,
- istnieje wiele rozwiązań, które można wykorzystać w celu ekstrakcji różnych sformułowań i potencjalnie można je wykorzystać w przypadku wdrożenia przygotowywanego rozwiązania do praktyki,
- podobne decyzje do ograniczania się do określonych typów fraz podejmowano w wielu istniejących pracach, np. [Hu i Liu 2004a; Kaji i Kitsuregawa 2007; Lu i in. 2011; Rill i in. 2012]. Jest to podyktowane tym, że przeważająca większość opinii wyrażana jest za pomocą niewielkiej liczby typów sekwencji.

Do implementacji reguł wykorzystano formalizm JAPE (Java Annotation Patterns Engine) [Cunningham, Maynard i Tablan 1999]. Przykład wykorzystanej reguły przedstawiono w załączniku G na stronie 201. Zaimplementowane reguły wzbogacono o dodatkowe warunki, które ograniczały zidentyfikowane sekwencje jedynie do tych, w których określone wyrazy z sekwencji miały zgodne ze sobą liczbę (pojedynczą lub mnogą), rodzaj i przypadek. Dodatkowo, na podstawie obserwacji z analizy korpusu, zaimplementowano reguły adnotujące zakresy tekstu będące negacjami lub wyrazami wskazującymi na to, że opinie wyrażane w danym zdaniu mogą mieć zmieniony sens (np. czasowniki takie jak „mógłby”, „powinien” , które zostały oznaczone przez adnotatorów podczas analizy korpusu jako zmieniające sens wyrażanych opinii). Fragmenty tekstu oznaczone przez reguły z tablicy 6.1, które znajdowały się w odległości do ośmiu wyrazów po takich adnotacjach sformułowań zmieniających sens opinii, nie były brane pod uwagę w dalszym przetwarzaniu.

Przetwarzanie korpusu z wykorzystaniem zidentyfikowanych reguł przeprowadzono za pomocą platformy GATE (General Architecture for Text Engineering)². Aby to było możliwe, konieczne było przetłumaczenie plików z adnotacjami zwracanych przez TaKIPI na format zrozumiały dla GATE, co przeprowadzono za pomocą specjalnie przygotowanego skryptu. Następnie, zaimple-

²gate.ac.uk

mentowano program w języku Jython, który pobierał zadane pliki i przetwarzał je za pomocą zdefiniowanego potoku przetwarzania GATE. Na potok składały się następujące moduły:

- *Simple tokenizer*, tj. moduł dzielący tekst na pojedyncze tokeny,
- *ANNIE gazetteer*, czyli moduł oznaczający w tekście wystąpienia słów z zadanych leksykonów, adnotujący słowa będące nazwami aspektów oraz słowa zmieniające sens wyrażanych opinii,
- *JAPE transducer* (przetwornik JAPE), czyli moduł przetwarzający teksty za pomocą zdefiniowanych reguł.

Wyekstrahowane frazy były następnie przefiltrowywane z wykorzystaniem tzw. stop listy. Odrzucane były te frazy, w których sformułowanie wykorzystane do oceny aspektu było na liście 500 najczęściej występujących wyrazów w języku polskim. Jest to częsta praktyka stosowana w takich sytuacjach, gdyż zakłada się, że najczęściej występujące słowa w danym języku często są funktorami, czyli nie niosą znaczenia istotnego w danym kontekście, a jedynie służą do konstrukcji bardziej złożonych zdań.

Z wykorzystaniem tak skonstruowanego potoku, w wyniku przetwarzania testowego korpusu, składającego się losowo wybranych 1600 recenzji tekstowych telefonów komórkowych (pochodzących z portalu cokupic.pl) wyekstrahowano 1514 fraz, z których na etapie filtrowania z wykorzystaniem stop listy odrzucono fraz 127.

Faza statystyczna

Ostatnią fazą w procesie ekstrakcji leksykonu sformułowań wyrażających opinie była statystyczna analiza kontekstu, w którym znajdowały się ekstrahowane frazy. Celem było odrzucenie tych przykładów, w których prawdopodobnie nie poddawano ocenie aspektów produktu. Każda wyekstrahowana fraza miała zostać przypisana do jednej z dwóch klas:

- zaakceptowane – frazy prawidłowo wyekstrahowane przez przygotowane reguły i które powinny zostać uwzględnione w dalszym przetwarzaniu,
- odrzucone – frazy, które nie będą dalej wykorzystywane w pracy.

Podejście to zostało zainspirowane rozwiązaniem stosowanym w systemach ujednoznaczniania podejść. Przykładowo, w pracach [Mihalcea, Banea i Wiebe 2007; Ng i Lee 1996] analizowano

m.in. jakie części mowy i jakie słowa statystycznie często pojawiały się w kontekście wystąpień danego słowa o różnych znaczeniach. Proponowana metoda ma być przeniesieniem tego podejścia na grunt ustalenia czy dana fraza służy wyrażaniu opinii, czy też nie.

W celu zdefiniowania tego, jakie cechy kontekstu mają być brane pod uwagę w trakcie klasyfikacji oraz zbudowania bazy pozytywnych i negatywnych przykładów, przeprowadzono szczegółową analizę fraz wyekstrahowanych w poprzedniej fazie. W tym celu ponownie dwóch adnotatorów miało za zadanie analizę 750 wyekstrahowanych automatycznie fraz. Mieli oni za zadanie, na podstawie sformułowanych wskazówek adnotacji (patrz załącznik D na stronie 189) przypisać każdą frazę do jednej z dwóch klas wspomnianych powyżej. Adnotatorzy byli ze sobą zgodni w 684 przypadkach, wśród których 579 przykładów było oznaczonych jako prawidłowo wyekstrahowanych, a 105 fraz oznaczono jako zidentyfikowane błędnie.

W trakcie dalszej analizy przykładów zidentyfikowano listę cech, które potencjalnie mogły rozróżnić przykłady prawidłowe od nieprawidłowych. Na podstawie wartości cech i przeprowadzonej nauki, klasyfikator miał określić, czy dane sformułowanie powinno zostać dodane do leksykonu, czy też nie. Takie cechy to m.in. liczba występujących w kontekście danej frazy poszczególnych części mowy, dłuższych fraz składających się z określonych części mowy i określonych słów. Kontekst zdefiniowano jako wszystkie wyrazy znajdujące się w oknie do trzydziestu znaków przed i po analizowanej frazie (kontekst ulegał skróceniu w przypadku wykrycia końca zdania). W sumie zdefiniowano 13 cech, do których należą:

- Liczba wystąpień w kontekście wyrazów o następujących tagach nadanych przez TaKIPI: *subst*, *adv*, *fin*, *pact*, *winien*, *qub* (w sumie 6 cech). Cechy te odpowiadają wybranym klasom gramatycznym słów występujących w kontekście frazy. Cechy takie wykorzystywano również w ujednoznacznianiu podejść we wspomnianych artykułach [Mihalcea, Banea i Wiebe 2007; Ng i Lee 1996].
- Liczba wyrazów pochodzących z następującej listy: „plus”, „minus”, „zaleta”, „wada”, „atut”, wraz z wszystkimi ich odmianami (1 cecha); wyrazy takie w analizowanych przykładach często wskazywały na to, że w zdaniu recenzent ocenia aspekty produktu lub usługi, a ich występowanie może wskazywać na to, że analizowana fraza została wyekstrahowana prawidłowo. Przykładem kontekstu, w którym występuje słowo z podanej listy jest „Plusem jest długa żywotność baterii”, gdzie klasyfikowaną frazą jest „długa żywotność baterii”. W tej frazie faktycznie wyrażana jest opinia o aspekcie recenzowanego produktu.

- Liczba wystąpień dłuższych fraz, identyfikowanych za pomocą specjalnie zdefiniowanych reguł. W trakcie analizy przykładów stwierdzono, że bardzo często recenzenci wyrażają w jednym zdaniu opinie o dużej liczbie aspektów. Przykładem takiego zdania jest „Świetna jakość dźwięku, żywotna bateria, dobry aparat foto i łatwa obsługa składają się na ten fantastyczny telefon”. Przyjmijmy, że klasyfikujemy frazę „dobry aparat foto”. W kontekście tej frazy znajdują się dodatkowe sformułowania identyfikowane przez reguły jako potencjalnie wyrażające opinie. W związku z tym, zdefiniowano dwie cechy, gdzie jedna z nich wymagała wystąpienia sekwencji identycznych ze zdefiniowanymi w tabeli 6.1 oraz druga, gdzie nazwy aspektu zastąpiono dowolnym rzeczownikiem (dla fraz, za pomocą których oceniane są aspekty nieuwzględnione w zdefiniowanej liście).
- Liczba zidentyfikowanych nazw aspektów (gdzie motywacja jest podobna do tej z poprzedniego punktu).
- Liczba wyrazów z ręcznie zdefiniowanej listy podstawowych wyrazów służących do wyrażania opinii, tj. „dobry”, „zły”, „świetny”, „kiepski”, „słaby”, „wspaniały”, „niezły”, „beznadziejny”; słowa te wskazują, że w danym kontekście recenzent prawdopodobnie wyraża opinię.
- Liczba przymiotników w stopniu wyższym; w trakcie analizy korpusu stwierdzono, że takie przymiotniki (np. „lepszy”, „szybszy”) często są stosowane w recenzjach porównawczych i nie wyrażają one opinii o produkcie, do którego przypisana jest recenzja. Recenzje porównawcze są jednak poza zakresem pracy (patrz sekcja 4.4).

Dla każdej frazy z wyekstrahowanej kolekcji, konstruowano cechy w opisany sposób. Dodatkowo, frazy te przypisane były przez adnotatorów do właściwych klas. Na tej podstawie, możliwe było przeprowadzenie uczenia maszynowego, którego celem było skonstruowanie klasyfikatora. W tym celu wykorzystano pakiet *scikit-learn*³, bibliotekę do uczenia maszynowego dla języka Python. Za pomocą przygotowanego skryptu przeprowadzono naukę naiwnego klasyfikatora Bayesa (patrz sekcja 3.3.1).

6.4.3 Ewaluacja rozwiązania

W tej sekcji zaprezentowana zostanie ewaluacja wypracowanego podejścia do ekstrakcji sformułowań służących do wyrażania opinii. Ewaluację przeprowadzono z wykorzystaniem miar precyzji, pełności i miary F_β (wyjaśnionych na stronie 93).

³<https://github.com/scikit-learn/scikit-learn>

Ze względu na fakt, że korpusy, którymi dysponowano, są dużych rozmiarów (składają się z kilkudziesięciu tysięcy recenzji), w trakcie ewaluacji znacznie większą wagę przywiązywano do precyzji niż do pełności. Jest to uzasadnione tym, że w tak dużych kolekcjach recenzji pominięcie pojedynczej frazy potencjalnie nie pociąga za sobą dużych kosztów, ponieważ jest duża szansa, że fraza ta może wystąpić w tej kolekcji wielokrotnie i ekstrakcja przynajmniej części z pozostałych wystąpień może przebiec pomyślnie.

Przeprowadzono ewaluację wyników osobno dla wyników fazy regułowej, opisanej w podsekcji 6.4.2 i fazy statystycznej, opisanej w podsekcji 6.4.2. Porównanie wyników tych dwóch faz miało wykazać, czy korzystając z zaproponowanej metody filtrowania fraz wykorzystującej uczenie maszynowe, można otrzymać rezultaty lepsze od tych będących wynikiem wyłącznie ekstrakcji regułowej. Jako główną miarę do porównania wyników wykorzystano miarę F_β . Jak wspomniano w sekcji 5.3.1, parametr β wpływa na to jak wysoki wpływ na miarę F_β mają precyzja i pełność; dla $\beta = 0.5$, precyzja ma dwa razy większy wpływ na tę miarę niż pełność. Ze względu na fakt, że, jak wspomniano, w trakcie ewaluacji znacznie większą wagę przywiązywano do precyzji, podjęto decyzję o ustaleniu tego parametru na poziomie 0,25, podobnie jak to miało miejsce w pracy [Haulrich 2009].

Ewaluację przeprowadzano z wykorzystaniem zaadnotowanych przykładów, będących wynikiem ręcznej adnotacji opisanej w sekcji 6.4.2 (traktowano je jako złoty standard). Analizowano 684 przykładów, które zostały zgodnie ocenione przez obu adnotatorów. Dla potrzeb porównania obu podejść przyjęto, że pełność w podejściu regułowym wynosiła 100%. Precyzję obliczono natomiast w ten sposób, że przykłady oznaczone jako „zaakceptowane” traktowano jako poprawnie wyekstrahowane, natomiast jako „odrzucone” jako wyekstrahowane błędnie.

Dla fazy statystycznej konieczne było podzielenie kolekcji przykładów na dwie części: część uczącą (na podstawie której uczono klasyfikator) oraz testową, którą najpierw poddawano klasyfikacji, a następnie obliczano precyzję i pełność tak, jak zaprezentowano w równaniach 6.1 i 6.2:

$$\text{precyzja} = \frac{p_{zz}}{p_{zz} + p_{oz}} \quad (6.1)$$

$$\text{pełność} = \frac{p_{zz}}{p_{zz} + p_{zo}} \quad (6.2)$$

gdzie:

- p_{zz} to przykłady przypisane do klasy „zaakceptowane” zarówno przez adnotatorów, jak i klasyfikator,
- p_{oz} to przykłady oznaczone przez adnotatorów jak „odrzucone” i przypisane przez klasyfikator do klasy „zaakceptowane”,
- p_{zo} to przykłady, które adnotatorzy oznaczyli jako prawidłowo wyekstrahowane („zaakceptowane”), ale odrzucone przez klasyfikator.

Podziału przykładów na kolekcje uczące i testowe dokonano z wykorzystaniem tzw. 10-krotnej krzyżowej walidacji. Polega ona na tym, że przykłady dzieli się na 10 części i eksperyment przeprowadzany jest dziesięciokrotnie, gdzie za każdym razem dziewięć części branych jest jako zestaw uczący, a jedna (za każdym razem inna) jako testowy. Następnie, otrzymane wyniki są uśredniane. Tym samym, testowanie odbywa się na każdym przykładzie z posiadanej kolekcji.

Faza	Precyzja	Pełność	$F_{0,25}$
ekstrakcja	84,6%	100%	85,4%
stop lista (filtr.)	90,6%	97,9%	91%
statystyczna	92,3%	83,5%	91,7%

Tabela 6.2: Ewaluacja wyników ekstrakcji fraz służących do wyrażania opinii. Źródło: opracowanie własne

Otrzymane rezultaty zostały zaprezentowane w tabeli 6.2. Przeprowadzone eksperymenty wykazały, że już korzystając z samej ekstrakcji regułowej otrzymane wyniki charakteryzowały się wysoką precyzją (ponownie, pełność na poziomie 100% została tu przyjęta umownie). Przefiltrowanie fraz za pomocą stop listy znacząco poprawiło precyzję, przy bardzo niewielkim spadku pełności. Natomiast zaproponowana w pracy dodatkowa faza filtrowania spowodowała dalszy wzrost precyzji o niemal 2 punkty procentowe, co przy tak wysokiej precyzji wyjściowej można uznać za satysfakcjonujący rezultat.

Podsumowując, wzbogacenie metody ekstrakcji regułowej o filtrowanie statystyczne na podstawie analizy kontekstu fraz (z wykorzystaniem uczenia maszynowego) pozwoliło na uzyskanie lepszych rezultatów według miary $F_{0,25}$. Zaproponowana metoda ekstrahuje frazy z precyzją na poziomie 92,3%. Jeśli natomiast chodzi o pełność opracowanego rozwiązania, to nie jest ona kluczowa ze względu na fakt, że dostępne korpusy są bardzo obszerne. Niweluje to negatywny

wpływ pomijania pojedynczych fraz ze względu na fakt, że większość par (*sformułowanie, aspekt*) występuje w korpusie wielokrotnie.

6.5 Identyfikacja polarności

Wykorzystując wypracowaną metodę ekstrakcji, możliwe było wygenerowanie słowników sformułowań używanych przez recenzentów do wyrażania opinii o różnych aspektach produktów i usług. W tym momencie, można było przystąpić do opracowania i implementacji metody realizującej główny cel pracy.

W tej sekcji opisano implementację wariantu metody, działającego dla korpusu recenzji telefonów komórkowych pobranych z portalu `cokupic.pl`. Jest to ten sam korpus, który wykorzystano w dotychczasowych eksperymentach opisanych w rozdziałach 5 i 6. Natomiast ewaluacja rozwiązań i wykazanie tezy zostanie zaprezentowane w rozdziale 7.

6.5.1 Wejście i wyjście dla metody

Wypracowana metoda ma na celu identyfikację polarności sformułowań służących do wyrażania opinii o aspektach analizowanych produktów z wykorzystaniem podsumowań opinii wielu recenzentów o tych produktach. Na portalu `cokupic.pl`, jak wcześniej wspomniano, recenzenci mogą wyrażać swoje opinie m.in. za pomocą list zalet i wad, przy czym listy te składają się z predefiniowanych opcji (patrz sekcja 4.3.2).

Na podstawie analizy opisanej powyżej w tym rozdziale przygotowano następujące dane wejściowe:

1. aspekt, którego dane sformułowanie dotyczy,
2. sformułowanie, za pomocą którego recenzenci oceniali dany aspekt,
3. listę recenzji, w których dane sformułowanie zostało użyte dla oceny aspektu, czyli tych recenzji, z których została wyekstrahowana dana para (*sformułowanie, aspekt*), w tym informację o tym, czy w danych recenzjach analizowany aspekt był oceniony w liście zalet i wad oraz, jeśli tak, to w jaki sposób,
4. listę produktów, które oceniane były w recenzjach wspomnianych w poprzednim punkcie (było to możliwe ze względu na fakt, że na portalu `cokupic.pl` recenzje były jednoznacznie przypisywane do produktów),

5. kolekcję podsumowań opinii wyrażonych w postaci list zalet i wad o wszystkich produktach wspomnianych w poprzednim punkcie, w tym podsumowań opinii dotyczących aspektu wspomnianego w punkcie pierwszym (o strukturze analogicznej do podsumowania zaprezentowanego w tabeli 4.1 na stronie 79).

Dla każdej pary (*sformułowanie*, *aspekt*) metoda dysponowała więc danymi, na podstawie których miała zostać podjęta decyzja o polarności tej pary. Do par przypisywano jedną z trzech polarności:

- pozytywna,
- negatywna,
- neutralna (brak opinii).

Połączenie polarności neutralnej z brakiem opinii można umotywić trudnością w ich rozgraniczeniu. Praktyka ta stosowana jest również w innych pracach dotyczących analizy wydźwięku [Liu 2011].

6.5.2 Przypisywanie polarności

Na podstawie omówionego wejścia, dla każdej pary generowano cechy (statystyki), które miały posłużyć do identyfikacji, jaka polarność powinna zostać do danej pary przypisana. Generowanie cech dla analizowanego scenariusza zostało szczegółowo opisane w sekcji 4.3.2. Dla otrzymanych wektorów cech, które są generowane dla każdej pary z leksykonu, następuje identyfikacja polarności. Zadanie to rozumiane jest w pracy jako problem klasyfikacji, gdzie każdą parę przypisuje się do jednej z trzech klas: pozytywnej, negatywnej i neutralnej (brak opinii).

Klasyfikacja przeprowadzana jest z wykorzystaniem uczenia maszynowego, a dokładniej naiwnego klasyfikatora Bayesa. Na podstawie ręcznie zaadnotowanych przykładów, tj. przykładów, do których manualnie przypisano właściwe polarności, przeprowadzana jest nauka modelu. Następnie, model jest wykorzystany do przypisywania polarności do kolejnych, nieprezentowanych wcześniej wektorów cech. Więcej informacji o procesie przygotowywania przykładów uczących dla posiadanych danych znajduje się w sekcji 7.1.

Ewaluacja zaprezentowanej metody znajduje się w kolejnym rozdziale pracy.

6.6 Podsumowanie

W rozdziale opisano sposób działania proponowanej metody oraz badania, jakie zostały przeprowadzone w trakcie jej wypracowywania. Ważnymi rezultatami tego rozdziału są szczegółowy opis wszystkich czynności, jakie muszą zostać przeprowadzone, aby możliwe była identyfikacja polarności dla aspektu, a także autorska metoda ekstrakcji do leksykonu sformułowań służących do wyrażania opinii, wykorzystująca do tego statystyczną analizę kontekstu sformułowań ekstrahowanych za pomocą zdefiniowanych reguł. Na tej podstawie, możliwe było przeprowadzenie eksperymentów oceniających jej skuteczność. Zostanie to opisane w kolejnym rozdziale pracy. Tym samym osiągnięte zostaną drugi i trzeci cel szczegółowy prezentowanej rozprawy.

Rozdział 7

Ewaluacja wypracowanej metody

Celem rozdziału jest omówienie wyników eksperymentów przeprowadzonych dla oceny skuteczności działania wypracowanej metody w porównaniu z wybranymi istniejącymi rozwiązaniami. Ewaluacja ma udowodnić tezę przyjętą dla pracy i sformułowaną we Wprowadzeniu, a także, razem z poprzednim rozdziałem, pozwolić na osiągnięcie drugiego i trzeciego celu szczegółowego rozprawy.

Wykazanie tezy zostanie przeprowadzone za pomocą trzech eksperymentów. Przeanalizowane zostanie wykorzystanie proponowanej metody w wariantach 1 i 2, opisanych w sekcjach 4.3.1 i 4.3.2. Przeprowadzone eksperymenty dotyczyły:

1. przypisywania polarności na podstawie podsumowań list zalet i wad – w tym eksperymencie sprawdzano, czy wypracowana metoda pozwala na uzyskanie lepszych (w porównaniu z wybranymi istniejącymi metodami) wyników identyfikacji polarności dla danych z portalu `cokupic.pl` i wariantu 2 z sekcji 4.3.2,
2. wykorzystania skonstruowanego leksykonu w analizie wydźwięku – eksperyment ten stanowi kontynuację eksperymentu pierwszego i ma na celu porównanie wyników analizy wydźwięku, przeprowadzonej z wykorzystaniem leksykonu otrzymanego za pomocą wypracowanej metody, z wynikami otrzymanymi przy korzystaniu z leksykonów wygenerowanych przez wybrane istniejące metody,
3. identyfikacji polarności na podstawie podsumowań ocen punktowych – za pomocą tego eksperymentu analizowano czy możliwe jest przeprowadzenie wnioskowania przedstawionego w pierwszym wariantcie metody, opisanym w sekcji 4.3.1, tj. identyfikowanie polarności wyłącznie na podstawie podsumowań opinii wielu recenzentów, którzy wyrażali swoje opinie

o aspektach restauracji za pomocą ocen punktowych; eksperyment został przeprowadzony z wykorzystaniem danych z portalu `gastronauci.pl`.

7.1 Eksperyment 1. Przypisywanie polarności na podstawie podsumowań list zalet i wad

W tej sekcji opisana zostanie ewaluacja wariantu w sekcji 4.3.2. Opracowana metoda miała przypisywać polarność (pozytywną, negatywną lub neutralną) do par (*sformułowanie, aspekt*) wyekstrahowanych z recenzji tekstowych, przy czym dysponowano łatwo przetwarzalnymi opiniami recenzentów, wyrażonymi w postaci listy zalet i wad z predefiniowanymi opcjami (opinie dotyczyły zdefiniowanych aspektów recenzowanych produktów i mogły być w stosunku do aspektu pozytywne, negatywne lub pominięte).

7.1.1 Dane

Eksperyment został przeprowadzony z wykorzystaniem recenzji telefonów komórkowych pobranych z portalu `cokupic.pl` (patrz sekcja 5.2.2). Do eksperymentu wybrano wszystkie recenzje telefonów opublikowane w jednym roku (2012). Powodem takiego wyboru recenzji był fakt, że możliwe jest, że pewne sformułowania mogą potencjalnie zmieniać swoją polarność (np. telefony o tym samym rozmiarze mogą być różnie oceniane w zależności od panującej mody). W przypadku analizy recenzji ze ściśle określonego przedziału czasowego, podobne trudności miały być zminimalizowane.

W ten sposób, do dalszych eksperymentów wybrano 3600 recenzji (wszystkie recenzje z 2012 roku). Następnie, na tych recenzjach przeprowadzono fazę ekstrakcji leksykonu za pomocą opracowanej metody opisanej w sekcji 6.4. W rezultacie, wyekstrahowano 2493 frazy, będące wystąpieniami 687 unikalnych par (*sformułowanie, aspekt*). Większość par miała bardzo niewielką liczbę wystąpień, w tym aż 421 par zostało wyekstrahowane tylko jeden raz. Jednocześnie, maksymalna liczba wystąpień pary w korpusie wynosiła 124. W tabeli 7.1 zaprezentowano 10 najczęściej występujących w analizowanym korpusie par wraz z liczbą ich wystąpień.

Wyekstrahowane pary zostały następnie ręcznie zaadnotowane. Adnotacja miała na celu zbudowanie bazy przykładów, w których do par (*sformułowanie, aspekt*) przypisane są polarności, t.j. jedna z trzech wartości:

- pozytywna – powinna być przypisywana do par, które użyte w zdaniu wyrażają pozytywną

aspekt	sformułowanie	wystąpienia
obsługa	prosty	124
obsługa	łatwy	109
aparat fotograficzny	dobry	96
żywołność baterii	słaby	75
rozmiar wyświetlacza	duży	43
żywołność baterii	dobry	41
cena	przystępny	36
rozdzielczość wyświetlacza	duży	35
cena	dobry	34
aparat fotograficzny	słaby	30

Tabela 7.1: Najczęściej ekstrahowane pary (*sformułowanie, aspekt*) (za pomocą metody opisanej w sekcji 6.4) w przyjętym korpusie recenzji. Źródło: opracowanie własne

opinię o danym aspekcie,

- negatywna – należy oznaczyć nią pary, które wykorzystane w zdaniu wyrażają negatywną opinię o aspekcie z danej pary,
- neutralna/brak opinii – ta wartość powinna być przypisana do par, w których sformułowanie nie wyraża ani pozytywnej, ani negatywnej opinii o aspekcie (opinia neutralna lub brak opinii).

Adnotacja została przeprowadzona przez dwóch niezależnych adnotatorów. Analizowali oni poszczególne wystąpienia danych par i na tej podstawie przypisywali do nich polarność. Dokładne reguły adnotacji zostały opisane w załączniku E na stronie 194. W procesie adnotacji, adnotatorzy mieli do dyspozycji cztery typy adnotacji, gdzie wyodrębniono opinię neutralną oraz brak opinii (brak opinii przypisywano do par, które pochodziły z błędnie wyekstrahowanych fraz), jednak w eksperymencie zostały one połączone, zgodnie z częstą praktyką stosowaną w literaturze [Liu 2011].

Wśród 697 par adnotatorzy nadali zgodne ze sobą adnotacje w 599 przypadkach, to jest dla 85,9% par. Dla pozostałych par polarność została ustalona przez tzw. sędziego, to jest jednego

z adnotatorów, który ponownie analizował kolejne niezgodne przypadki i podejmował decyzję, jaką polarność ostatecznie do danej pary należy przypisać. W ten sposób otrzymano tzw. złoty standard, w którym dla 347 par nadano polarność pozytywną, dla 106 negatywną, a 244 par uznano jako neutralne lub niewyrażające opinii.

Dla każdej pary (*sformułowanie, aspekt*) generowano wektory cech zgodnie z procesem przedstawionym na rysunku 4.4 na stronie 82.

7.1.2 Wypracowana metoda

Na podstawie wektorów cech, metoda miała automatycznie przypisywać parom odpowiednie wartości (pozytywna, negatywna, neutralna / brak opinii). Przypisywanie wspomnianych wartości do par rozumiano jako problem klasyfikacji, tj. każdą parę przypisywano do jednej z trzech klas. Tak rozumiany problem można rozwiązać za pomocą istniejących metod uczenia maszynowego, np. z wykorzystaniem naiwnego klasyfikatora Bayesa lub wektorów maszyn wspierających (metody te zostały opisane w sekcji 3.3.1). Metody uczenia maszynowego konstruują odpowiedni model na podstawie dostarczonych przykładów (zazwyczaj przygotowanych ręcznie), w których obiekty o przypisanych wektorach cech przypisane są do różnych klas. Następnie, model taki może być wykorzystany do przypisania nowych, nieznanymi wcześniej systemowi obiektów do odpowiednich klas (na podstawie ich wektorów cech). Do konstruowania modelu i klasyfikacji wykorzystano wspomniany już moduł do języka programowania Python o nazwie *scikit-learn*¹.

W przypadku opisywanej metody, dysponowano przykładami uczącymi dzięki przygotowanym adnotacjom, opisanym w sekcji 7.1.1. Wejściem do uczenia maszynowego były więc wektory cech wraz z adnotacjami, interpretowanymi tu jako informacja o tym, do jakiej klasy dany przykład należy zaliczyć. Na tej podstawie, w sposób automatyczny konstruowany był model naiwnego klasyfikatora Bayesa, za pomocą którego możliwe było przypisywanie nowych par do tych samych klas. W zależności od tego, do której klasy przykład został zaklasyfikowany, przypisywano mu polarność pozytywną, polarność negatywną lub stwierdzano, że wyraża on opinię neutralną lub brak opinii.

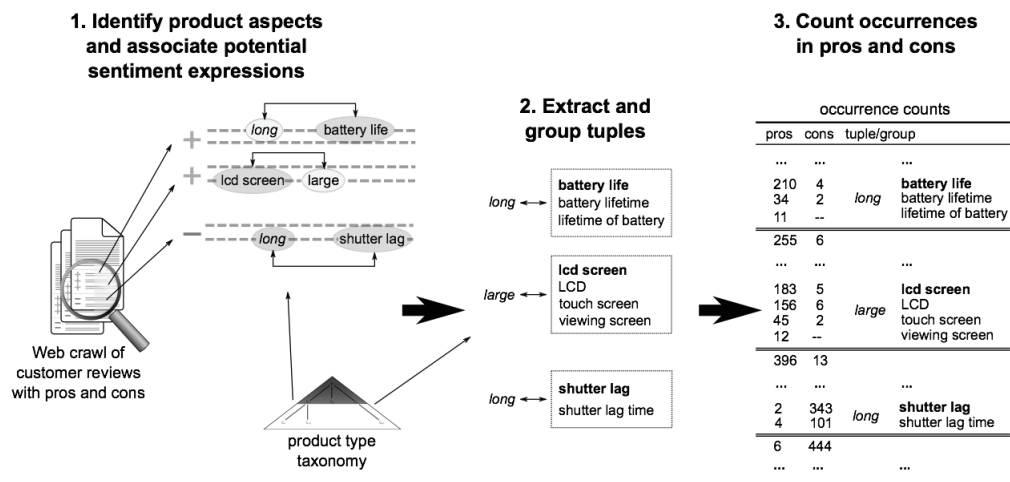
7.1.3 Metody bazowe

W eksperymencie porównano poprawność działania zaproponowanej metody z dwiema metodami bazowymi, wobec których opracowana metoda miała wykazać korzyści z wykorzystania podsumo-

¹<http://scikit-learn.org/stable/>

wań wielu opinii o produktach i usługach w procesie identyfikacji polarności. Każda z opisywanych metod została przetestowana dla tych samych danych, opisanych w podsekcji 7.1.1.

Pierwsze z podejść było analogiczne do zaproponowanego w prezentowanej pracy, jednak nie wykorzystywało informacji o zbiorczych opiniach recenzentów o produktach i usługach. Metoda ta odpowiadała podejściu przedstawionemu na rysunku 4.4, przy czym kończyła swe działanie na obliczeniu cech wyłącznie na podstawie przetwarzalnych opinii o recenzowanych produktach wyrażonych przez autorów recenzji, w których dana para występowała. Metoda ta obejmowała więc dane o parach pochodzące na rysunku 4.4 wyłącznie z szarych pól. Metoda ta dalej będzie zwana metodą bazową 1.



Rysunek 7.1: Proces ekstrakcji leksykonu i przygotowania danych do identyfikacji polarności w pracy Jürgena Broña, skąd zaczerpnięto metodę wykorzystaną w eksperymentach jako jedną z metod bazowych. Źródło: [Broń 2013]

Kolejną metodą, do której porównane zostanie wypracowane podejście, jest metoda wypracowana przez Jürgena Broña w jego pracy doktorskiej opublikowanej w 2013 roku [Broń 2013], w której przeprowadzał on badania nad analizą wydźwięku dla tekstów napisanych w języku angielskim. Będziemy określać ją mianem metody bazowej 2. Metoda identyfikacji polarności w omawianej pracy działała na danych podobnych do wykorzystywanych w rozprawie, ale pożytkiwanych w inny sposób. W związku z tym, należy dokładnie wyjaśnić w jaki sposób działa wspomniana metoda i w jaki sposób porównano jej wyniki z metodą proponowaną w prezentowanej pracy.

Jak przedstawiono na rysunku 7.1, w metodzie bazowej 2, pary (*sformułowanie, aspekt*) eks-

trahowane były z opinii wyrażanych wyłącznie w postaci list zalet i wad, przy czym poszczególne pozycje z tych list zalet i wad były wprowadzane przez użytkownika. Innymi słowy, w listach zalet i wad znajdowały się dowolne krótkie wypowiedzi tekstowe, a nie, jak w naszym przypadku, predefiniowane opcje. Jednocześnie, jako kolekcję strukturyzowaną, na podstawie której w tej metodzie przeprowadzano nadzorowanie odległe, była informacja o tym, ile razy dana para została wyekstrahowana z wypowiedzi zawartej w liście zalet, a ile razy w liście wad.

Przypomnijmy, że eksperymenty przeprowadzane są na korpusie recenzji, gdzie zalety i wady są predefiniowane, co umożliwia traktowanie ich jako strukturyzowanej kolekcji opisującej kolekcję niestrukturyzowaną, którą są recenzje wyrażone w postaci wypowiedzi tekstowych. Oba korpusy różnią się więc między sobą, jednak z punktu widzenia nadzorowania odległego przetwarzanie odbywa się analogicznie. Aby to dokładniej zrozumieć wystarczy zauważyć, że na rysunku 7.1, w tabeli z prawej strony, poszczególne wiersze mają analogiczną strukturę do części wektora cech przedstawionego na rysunku 4.4 (przy czym na rysunku 7.1 wartości nie są znormalizowane). Korpus analizowany w tym rozdziale jest bardziej wymagający od tego wykorzystanego w pracy [Broś 2013], gdyż, jak wielokrotnie wspomniano, w wielu przypadkach opiniom wyrażanym w wypowiedziach tekstowych nie towarzyszą w tej samej recenzji odpowiednie pozycje w liście zalet i wad.

Na podstawie tak pozyskanych danych (zaprezentowanych na rysunku 7.1 w tabeli z prawej strony), metoda bazowa 2 ustalała polarność poszczególnych par na podstawie statystycznego testu stosunku wiarygodności, za pomocą którego ustalano prawdopodobieństwo wystąpienia danej pary na liście zalet i liście wad (w naszym przypadku w wypowiedzi tekstowej, do której w przypisanej liście zalet i wad dany aspekt był oceniony pozytywnie lub negatywnie).

W trakcie eksperymentu odtworzono testy dokładnie w sposób opisany w pracy [Broś 2013]. Hipoteza zerowa głosiła, że prawdopodobieństwo tego, że dana para pojawia się w wypowiedziach tekstowych, dla których w liście zalet i wad dany aspekt był oceniony jako zaleta oraz tego, że w liście był oceniony jako wada, są sobie równe. W przypadku braku możliwości odrzucenia hipotezy zerowej, daną parę klasyfikowano jako wyrażającą opinię neutralną lub niewyrażającą opinii (neutralna / brak opinii). W przeciwnym przypadku, przypisywano jej polarność na podstawie tego, które prawdopodobieństwo było wyższe.

7.1.4 Procedura przeprowadzenia eksperymentu

Przeprowadzony eksperyment miał na celu porównanie skuteczności metody proponowanej w rozprawie z metodami bazowymi. Został on przeprowadzony z wykorzystaniem adnotacji opisanych w sekcji 7.1.1. W przypadku metody bazowej 2 (czyli metody przedstawionej w pracy [Broń 2013]), eksperyment został przeprowadzony na wszystkich posiadanych danych. Jednocześnie, w przypadku metody bazowej 1 i metody proponowanej w rozprawie, część danych musiała zostać wykorzystana jako przykłady w trakcie uczenia maszynowego.

Aby umożliwić wykorzystanie wszystkich adnotacji do przetestowania metody i zapewnić porównywalność wyników uzyskanych dla wszystkich testowanych metod, dla testowania metody bazowej 1 i metody proponowanej w rozprawie zastosowano tzw. dziesięciokrotną krzyżową walidację, opisaną już na stronie 128.

Dla potrzeb oceny testowanych metod wykorzystano standardowe miary precyzji, pełności i miary F_1 w rozbiciu na poszczególne klasy, do których przypisywane były pary, czyli polarność pozytywna, polarność negatywna i opinia neutralna/brak opinii. Dodatkowo, obliczono dwie miary zbiorcze, mające podsumować rezultaty otrzymane dla poszczególnych klas. Były to średnie ważone miar F_1 , gdzie wagami były liczebności przykładów z danych klas. Wykorzystano tu dwie miary:

- $\overline{F_1}$ jest to średnia ważona miar F_1 dla wszystkich klas (pozytywna, negatywna, neutralna/brak opinii); miara ta bierze pod uwagę wszystkie klasy,
- $\overline{F_{1PN}}$ to średnia ważona miar F_1 tylko dla dwóch klas: polarność pozytywna i polarność negatywna (nie jest brana pod uwagę klasa neutralna/brak opinii). Wyodrębniono te klasy ze względu na fakt, że tylko one mogą być wykorzystane w procesie analizy wydźwięku. Jak wspomniano, sformułowania wyrażające opinię neutralną i te niewyrażające opinii w praktyce często są traktowane tak samo [Liu 2011] i w związku z tym nie bierze się ich w trakcie analizy wydźwięku pod uwagę. Miara ta ma więc wskazać, która metoda pozwoli otrzymać najlepsze rezultaty w procesie analizy wydźwięku.

Podsumowując, miara $\overline{F_{1PN}}$ wskazuje, z jaką skutecznością dany leksykon może być wykorzystany w analizie wydźwięku i trafniej przewiduje, który leksykon pozwoli na uzyskanie lepszych rezultatów w tym procesie. Natomiast miara $\overline{F_1}$ uwzględnia dodatkowo, na ile trafnie dana metoda identyfikuje pary, które nie powinny być wykorzystywane w analizie wydźwięku.

Statystyki obliczono w dwóch wariantach. Pierwszy z nich traktował każdą parę z równą wagą niezależnie od tego, ile razy występowała ona w analizowanym korpusie (tj. ile razy pojawiła się w recenzjach wykorzystanych na etapie ekstrakcji). Natomiast drugi wariant przypisywał wszystkim parom wagi równe temu, ile razy dana para wystąpiła w korpusie. Wariant pierwszy potencjalnie miał dawać niższe statystyki ze względu na fakt, że pary o niewielkiej liczbie wystąpień prawdopodobnie trudniej przypisać do właściwych klas (gdyż często nie posiadamy o nich wystarczającej ilości danych). Wariant drugi dawał rezultaty pozwalające lepiej przewidywać skuteczność otrzymywanych leksykonów w procesie analizy wydźwięku, ponieważ do par częściej występujących przypisywano wyższe wagi. Należy oczekiwać, że w innym korpusie częstości występowania poszczególnych par byłyby analogiczne. W związku z tym, trafność przypisania właściwej polarności do częściej występujących par sprawi, że w większej liczbie przypadków na etapie analizy wydźwięku opinia recenzenta zostanie odgadnięta prawidłowo.

7.1.5 Uzyskane wyniki i ich interpretacja

Rezultaty otrzymane w eksperymencie zaprezentowano w tabelach 7.2 (wariant 1 liczenia statystyk) i 7.3 (wariant 2). Rezultaty są o tyle interesujące, że wyraźnie pokazują, że poszczególne metody mają różną skuteczność dla różnych miar i klas (polarności).

Na początku, warto porównać metodę prezentowaną w rozprawie z metodą bazową 1. Należy przypomnieć, że metody te są do siebie bardzo podobne i różnica pomiędzy nimi polega na tym, że w metodzie prezentowanej w rozprawie wzbogacono metodę bazową o informacje o podsumowaniach wielu opinii o produktach i ich aspektach, w recenzjach których wykorzystywano dane pary (*sformułowanie, aspekt*). Prezentowana metoda działała lepiej od metody bazowej dla obu miar $\overline{F_1}$ i $\overline{F_{1PN}}$, przy czym dla drugiej z nich wzrost skuteczności był niewielki (dla obu wariantów liczenia statystyk wynosił on mniej niż 1%), natomiast dla pierwszej był już znacząco większy. Wykazuje to prawdziwość tezy, gdyż wykorzystanie podsumowań faktycznie pozwoliło na uzyskanie lepszych rezultatów.

Porównanie prezentowanej metody z metodą bazową 2 jest także na korzyść metody prezentowanej w pracy. Według miary $\overline{F_1}$ obie metody mają praktycznie identyczną skuteczność, natomiast ich porównanie za pomocą miary $\overline{F_{1PN}}$ wypada na korzyść metody proponowanej w rozprawie. Jak wspomniano, z punktu widzenia wykorzystania wyników działania metody (otrzymanego leksykonu z przypisanymi polarnościami) w analizie wydźwięku, ważniejsza jest miara $\overline{F_{1PN}}$. W związku z tym eksperyment wykazał, że zaproponowana metoda może pozwolić uzyskać

polarność	statystyka	bazowa 1	bazowa 2	prezentowana w rozprawie
pozytywna	precyzja	61,23%	77,27%	66,25%
	pełność	97,35%	55%	85,43%
	F_1	74,99%	64,26%	74,4%
negatywna	precyzja	65,81%	66,15%	56,66%
	pełność	41,77%	41,75%	51,45%
	F_1	49,71%	51,19%	52,65%
neutralna / brak opinii	precyzja	15,27%	36,07%	40,18%
	pełność	1,16%	65,09%	18,84%
	F_1	2,13%	46,41%	24,94%
wszystkie	$\overline{F_1}$	50,69%	57,15%	57,14%
	$\overline{F_{1PN}}$	69,11%	61,22%	69,34%

Tabela 7.2: Statystyki precyzji, pełności i miary F dla poszczególnych polarności dla trzech porównywanych metod bez uwzględnienia częstości występowania poszczególnych par w korpusie. Źródło: opracowanie własne

lepsze wyniki w stosunku do tej metody bazowej.

Warto dokładniej przeanalizować dane w celu ustalenia przyczyn, dla których występowały różnice w skuteczności różnych metod oraz pomiędzy różnymi klasami.

W przypadku porównania proponowanej metody z metodą bazową 1 należy zauważyć, że przewagi proponowanej metody można upatrywać w prawidłowym klasyfikowaniu par o niewielkiej liczbie wystąpień. Jest to spowodowane tym, że dla par o większej liczbie obserwacji bardziej prawdopodobne jest, że w którejś z recenzji, w których dana para wystąpiła, została również wyrażona opinia o danym aspekcie w liście zalet i wad. W posiadanym korpusie faktycznie wystąpiła taka zależność, co zaprezentowano w tabeli 7.4. Jeśli dany aspekt nie był oceniany w liście zalet i wad w żadnej recenzji, w której dana para wystąpiła, cechy 1 i 2 w wektorze cech zgodnym z rysunkiem 4.4 miały zerowe wartości. W takim przypadku, metoda bazowa 1 nie miała żadnych danych, które mogłaby wykorzystać do właściwego przypisania pary do klasy. Natomiast metoda proponowana w rozprawie, pomimo tego, że miała zerowe wartości cech 1 i 2, mogła

polarność	statystyka	bazowa 1	bazowa 2	prezentowana w rozprawie
pozytywna	precyzja	79,24%	87,15%	82,53%
	pełność	98,65	64,86%	94,91%
	F_1	87,61	74,89%	88,07%
negatywna	precyzja	76,14%	85,56%	69,3%
	pełność	58,05%	69,37%	63,95%
	F_1	63,84%	76,62%	64,85%
neutralna / brak opinii	precyzja	15,91%	20,11%	38,51%
	pełność	3,12%	50,52%	15,41%
	F_1	4,87%	28,77%	20,69%
wszystkie	$\overline{F_1}$	60,85%	62,21%	65,67%
	$\overline{F_{1PN}}$	82,08%	74,89%	82,67%

Tabela 7.3: Statystyki precyzji, pełności i miary F dla poszczególnych polarności dla trzech porównywanych metod z uwzględnieniem częstości występowania poszczególnych par w korpusie. Źródło: opracowanie własne

wykorzystać informacje o tym, jakie były zbiorcze opinie wielu recenzentów o aspektach produktów ocenianych w recenzjach, w których dane pary występowały. Sytuacje, w których wartości odpowiadające podsumowaniom wielu opinii w wektorze również były zerowe (były to cechy 3 i 4 zgodnie z rysunkiem 4.4) były tu znacznie rzadsze. Było to w sumie tylko 48 przypadków (45 dla par z pojedynczym wystąpieniem i 3 dla par o dwóch wystąpieniach). Stąd, dla par o niewielkiej liczbie wystąpień, obie metody dawały znacznie różniące się rezultaty. Natomiast dla par z większą liczbą wystąpień, różnice pomiędzy metodami były coraz mniejsze, co zaprezentowano w tabeli 7.5.

Również w przypadku metody bazowej 2, jej zachowanie się dla par o niewielkiej liczbie wystąpień ma decydujący wpływ na uzyskane statystyki. Główną siłą tej metody, zgodnie z danymi zaprezentowanymi w tabeli 7.2, była bardzo wysoka pełność otrzymana dla kategorii neutralna/brak opinii, przy bardzo małych wartościach tej statystyki uzyskanych dla pozostałych dwóch metod. Ponownie, jest to bezpośrednio spowodowane zachowaniem się poszczególnych

wystąpienia	dane dostępne	brakuje danych
1	127	246
2	67	30
3	26	7
4	18	4
5	14	2
6	11	1
7	9	0
...

Tabela 7.4: Liczebności par o danej liczbie wystąpień w analizowanym korpusie, dla których metody bazowe nie posiadały przesłanek do przypisywania polarności do par (*sformułowanie, aspekt*). Wartości w kolumnie „dane dostępne” odpowiadają przypadkom, gdy cecha pierwsza lub druga w wektorze cech zgodnym z rysunkiem 4.4 miały niezerową wartość, w związku z czym istniały przesłanki do identyfikacji polarności dla metod bazowych pierwszej i drugiej. Natomiast dla przypadków, którym odpowiadają wartości z kolumny „brakuje danych”, obie cechy miały zerowe wartości, a tym samym metody bazowe nie posiadały takich przesłanek. Źródło: opracowanie własne.

metod w sytuacji braku danych w posiadanym wektorze cech. Metoda bazowa 2, działająca na zasadzie testu statystycznego, nie posiadała tu informacji niezbędnych do odrzucenia hipotezy o równym prawdopodobieństwie występowania danej pary w recenzjach, w których dany aspekt był oceniany jako zaleta i w których był on oceniany jako wada. W związku z tym, wszystkie takie obserwacje (a było ich aż 290) były klasyfikowane jako neutralne/brak opinii. Przypisanie tak dużej liczby obserwacji do tej klasy skutkowało właśnie wysokim podbiciem pełności (przy czym precyzja przypisywania par do tej klasy dla metody bazowej 2 była niższa niż dla metody proponowanej w rozprawie).

Analogicznie, bardzo wysoka pełność metody bazowej 1 otrzymana dla klasy odpowiadającej polarności pozytywnej wynika dokładnie z tego samego powodu. Model wygenerowany przez uczenie maszynowe, w przypadku napotkania pary, dla której cechy 1 i 2 w wektorze cech były zerowe, zawsze przypisywał takie pary do klasy pozytywna, co pozwoliło na uzyskanie bardzo

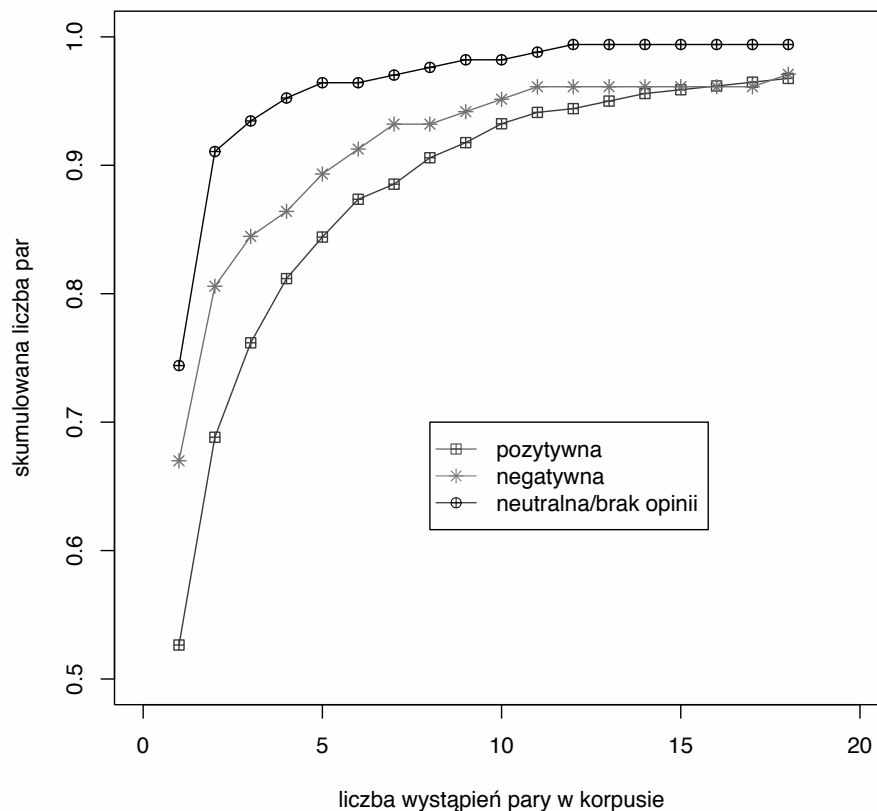
wysokiej pełności, przy jednoczesnym pogorszeniu precyzji. Z tego też powodu, pełność dla klasy neutralna/brak opinii była dla tej metody katastrofalna (poniżej 2%).

Wystąpienie	Liczba przypisań	
	niezgodnych	zgodnych
1	92	281
2	8	89
3	2	31
4	1	21
...

Tabela 7.5: Liczba niezgodnych klasyfikacji między metodą bazową 1 a metodą proponowaną w rozprawie w zależności od liczby wystąpień par. Źródło: opracowanie własne

Metoda zaproponowana w rozprawie, dzięki uwzględnieniu dodatkowych danych, mogła podczas przypisywania par z zerowymi cechami 1 i 2 wykorzystać wartości pozostałych cech. Dzięki temu, nie przypisywała ślepo wszystkich wartości do określonej klasy. Proponowana metoda, na podstawie posiadanych podsumowań opinii wielu recenzentów o produktach, dla danych z brakującymi cechami 1 i 2 przypisała 190 par do klasy pozytywna, 24 do klasy negatywna i 76 do klasy neutralna/brak opinii. Dzięki temu metoda ta uzyskała najwyższą precyzję w klasie neutralna/brak opinii oraz poprawę precyzji wobec metody bazowej 1 w klasie odpowiadającej pozytywnej polarności.

Warto również przeanalizować powód, dla którego analizowane metody znacznie gorzej rozpoznają pary o polarności negatywnej i pary neutralne, niż te o polarności pozytywnej. Fakt ten można wytłumaczyć tym, że w klasach negatywnej i neutralna/brak opinii jest stosunkowo znacznie więcej par o niewielkiej liczbie wystąpień, niż ma to miejsce dla klasy odpowiadającej polarności pozytywnej. Zostało to zobrazowane na rysunku 7.2, na którym widać m.in., że dla klasy negatywnej, liczba par o tylko jednym wystąpieniu stanowiła ponad 65% wszystkich par, dla klasy neutralnej/brak opinii ponad 70%, natomiast dla klasy pozytywnej było to tylko niewiele ponad 50%. Dodatkowo, nierówne liczebności poszczególnych klas (klasa negatywna jest ponad trzykrotnie mniej liczna niż pozytywna) również mogą mieć wpływ na prawidłowość przypisywania par do klas przez metody wykorzystujące uczenie maszynowe, tj. metodę bazową 1 i metodę proponowaną w rozprawie. Powodem może być tu niewielka liczba obserwacji z tych klas, które



Rysunek 7.2: Rozkład liczby par (*sformułowanie, aspekt*) o liczbie wystąpień nie większej niż dana wartość w rozbiciu na pary przypisane przez adnotatorów do różnych klas. Źródło: opracowanie własne

system ma do dyspozycji na etapie nauki.

7.2 Eksperyment 2. Wykorzystanie skonstruowanego leksykonu w analizie wydźwięku

Celem eksperymentu było określenie z jaką trafnością możliwe jest przeprowadzenie analizy wydźwięku z wykorzystaniem leksykonów uzyskanych w wyniku zastosowania metod opisanych w sekcji 7.1. Jako dane wejściowe do eksperymentu posiadano trzy leksykony, gdzie każdy został wygenerowany z wykorzystaniem jednej z analizowanych metod. Dla metod wykorzystujących uczenie maszynowe, leksykon generowano stopniowo, w trakcie przeprowadzania dziesięciokrotnej

krzyżowej walidacji. Dla każdego z otrzymanych leksykonów, na jego podstawie, dla posiadanej kolekcji fraz usiłowano automatycznie zidentyfikować opinię, jaką recenzent wyrażał o pewnym aspekcie, którego nazwa pojawiła się w danej frazie.

7.2.1 Przebieg eksperymentu i wykorzystane dane

Eksperyment został przeprowadzony na korpusie recenzji telefonów komórkowych z portalu `cokupic.pl`. W trakcie eksperymentu wykorzystano inny korpus niż w poprzednim eksperymencie, ze względu na fakt, że wówczas analiza wydźwięku byłaby przeprowadzona na tych samych danych, z wykorzystaniem których wygenerowano leksykon sformułowań służących do wyrażania opinii, co prowadziło do przekładań w uzyskanych wynikach. Pobrano więc wszystkie recenzje z tego portalu z roku 2011 (było ich w sumie 6909) i przeprowadzono na nich ekstrakcję regułową według opisu w sekcji 6.4.2, tj. przeprowadzono ekstrakcję regułową połączoną z filtrowaniem z wykorzystaniem stop listy i fazą filtrowania z wykorzystaniem uczenia maszynowego, dokładnie jak opisano to w poprzednim rozdziale. W ten sposób otrzymano 5462 fraz. Następnie odrzucono te frazy, w których występowały pary (*sformułowanie, aspekt*), których nie było w posiadanych leksykonach. W rezultacie otrzymano 4300 fraz.

W tym miejscu, wprowadzono dodatkowe ograniczenie odnośnie danych, na których eksperyment miał być przeprowadzony. Jak wykazano w sekcji 7.1.5, różnice pomiędzy poszczególnymi leksykonami miały miejsce przede wszystkim w podzbiorze par, które miały niewielką liczbę wystąpień w analizowanym korpusie. W związku z tym, aby dokładniej przeanalizować różnice działania analizowanych metod w tym podzbiorze, zawężono dane wykorzystywane w eksperymencie do fraz, gdzie występowały takie właśnie rzadko występujące pary. Jako rzadko występujące przyjęto pary występujące mniej niż 5 razy. Po odrzuceniu fraz z częściej występującymi parami posiadana kolekcja liczyła 1021 fraz. Spośród nich losowo wybrano 472 frazy i na nich przeprowadzono eksperyment.

Do wyekstrahowanych fraz przypisywano wydźwięk z wykorzystaniem posiadanych leksykonów. Każda fraza miała przypisaną parę (*sformułowanie, aspekt*), na podstawie której dana fraza została wyekstrahowana. Wydźwięk przypisywano poprzez wyszukiwanie w odpowiednim leksykonie danej pary i określanie, jaka polarność jest do niej w leksykonie przypisana. Jak zaznaczono w sekcji 6.4.2, dla potrzeb eksperymentu odrzucano wszystkie frazy, w których potencjalnie mógł być zmieniony sens danego sformułowania (czyli takie, gdzie np. występowała negacja w niewielkiej odległości przed daną parą), w związku z czym nie było konieczne przeprowadzanie

dodatkowych wnioskowań.

Poprawność analizy wydźwięku przeprowadzonej z wykorzystaniem poszczególnych leksykonów określono poprzez porównanie otrzymanych wyników z tym, w jaki sposób analogiczne zadanie zostało wykonane przez ludzi. W tym celu, posiadany korpus fraz został zaadnotowany przez dwóch ekspertów, którzy przypisywali do każdej frazy odpowiednią opinię, zgodnie ze wskazówkami adnotacji zawartymi w załączniku F na stronie 198. Adnotatorzy zgodnie ocenili około 88% fraz, zaś pozostałe frazy zostały ponownie przeanalizowane przez jednego z adnotatorów, który przeprowadzał dodatkową analizę i na tej podstawie przypisywał finalną opinię. W ten sposób uzyskano finalną adnotację, w której do 332 fraz przypisano opinię pozytywną, do 60 negatywną, a 80 uznano jako wyrażające opinię neutralną lub niewyrażające żadnej opinii o danym aspekcie.

7.2.2 Otrzymane wyniki i ich interpretacja

Wyniki ewaluacji analizy wydźwięku z wykorzystaniem leksykonów otrzymanych na podstawie poszczególnych analizowanych metod zaprezentowano w tabeli 7.6 na stronie 154. Ponownie wykorzystano statystyki precyzji, pełności i miary F dla poszczególnych klas, do których mogły być przypisywane frazy. Zaprezentowano również wartości statystyki F uśrednione dla wszystkich klas ($\overline{F_1}$) oraz tylko dla klas pozytywna i negatywna ($\overline{F_{1PN}}$) z wagami odpowiadającymi liczebności poszczególnych klas.

Otrzymane wyniki, co do rzędu wielkości poszczególnych statystyk, zgadzały się z rezultatami zaprezentowanymi w tabeli 7.3. Najlepsze wyniki, zarówno zgodnie z miarą $\overline{F_1}$, jak i $\overline{F_{1PN}}$, uzyskano za pomocą metody proponowanej w rozprawie. Na podstawie analizy uzyskanych danych stwierdzono, że główną siłą prezentowanej metody był korzystny stosunek precyzji do pełności. Nie we wszystkich statystykach proponowana metoda miała najwyższe wartości, jednak w takich przypadkach zawsze były one niewiele gorsze od metody, dla której wartość danej statystyki była najwyższa. Aby to dokładniej ukazać, obliczono wariant miary F z wartością β ustaloną na 0,5 ($F_{0,5}$), która kładzie dwukrotnie większy nacisk na precyzję od miary F_1 . Miara ta powinna być stosowana w sytuacjach, gdy kluczowe jest uzyskanie wysokiej precyzji, przy jednoczesnym akceptowalnym poziomie pełności. Otrzymane rezultaty wykazały, że w takiej sytuacji metoda proponowana w rozprawie byłaby bezkonkurencyjna i otrzymałaby najwyższą wartość dla każdej z klas.

Przeprowadzony eksperyment wykazał, że leksykon wygenerowany za pomocą proponowanej metody pozwala na otrzymanie dokładniejszych wyników analizy wydźwięku niż przyjęte metody

bazowe w przypadku fraz o niewielkiej liczbie wystąpień, dla których istniejące metody często nie posiadają wystarczającej ilości danych do właściwego przeprowadzenia wnioskowań.

7.3 Eksperyment 3. Identyfikacja polarności na podstawie podsumowań ocen punktowych

W ostatnim eksperymencie przeanalizowano wariant pierwszy metody, opisany w sekcji 4.3.1. Zgodnie z tym wariantem, system miał uczyć się polarności sformułowań napotykaných w recenzjach określonych dóbr nawet wówczas, gdy wypowiedzi tekstowej, w której wystąpiło dane sformułowanie, nie towarzyszyła żadna opinia wyrażona przez recenzenta w przetwarzalnym formacie. Nauka odbywała się tu z wykorzystaniem podsumowania opinii o recenzowanym bycie, pobranego z innego źródła, np. zdefiniowanego portalu z recenzjami dóbr z danej kategorii. Takie źródło będziemy dalej nazywać źródłem referencyjnym.

7.3.1 Identyfikacja restauracji

Podstawowym wyzwaniem w proponowanym scenariuszu, wychodzącym jednak poza obszar zainteresowania prezentowanej pracy, jest identyfikacja tego, którego dobra dotyczy dana recenzja oraz ujednoznacznianie nazw tych dóbr. Przykładowo, posiadamy wypowiedź tekstową dotyczącą pewnej restauracji o określonej nazwie i chcemy pozyskać podsumowanie opinii o tej restauracji z posiadanej przez nas kolekcji podsumowań opinii ze źródła referencyjnego. W tym momencie może przykładowo wystąpić sytuacja, gdy w referencyjnym źródle restauracja ta figuruje pod inną nazwą. Znalezienie właściwej restauracji może wymagać przeprowadzenia dodatkowych wnioskowań.

Opisana sytuacja wchodzi w zakres problemu ujednoznaczniania nazw encji, często stosowanego choćby dla celów wykrywania zduplikowanych wpisów w bazach danych [Bilenko i Mooney 2003; Cohen i in. 2003]. Dla domeny restauracji, będącej obszarem zainteresowania w eksperymencie, problem ten był analizowany przykładowo w pracy [Łazaruk i in. 2012]. Jak jednak zaznaczono, zagadnienie to znajduje się poza obszarem zainteresowania pracy i w tym rozdziale przyjmujemy, że dysponujemy metodą będącą w stanie wybierać właściwe podsumowania ze źródła referencyjnego z wysoką precyzją.

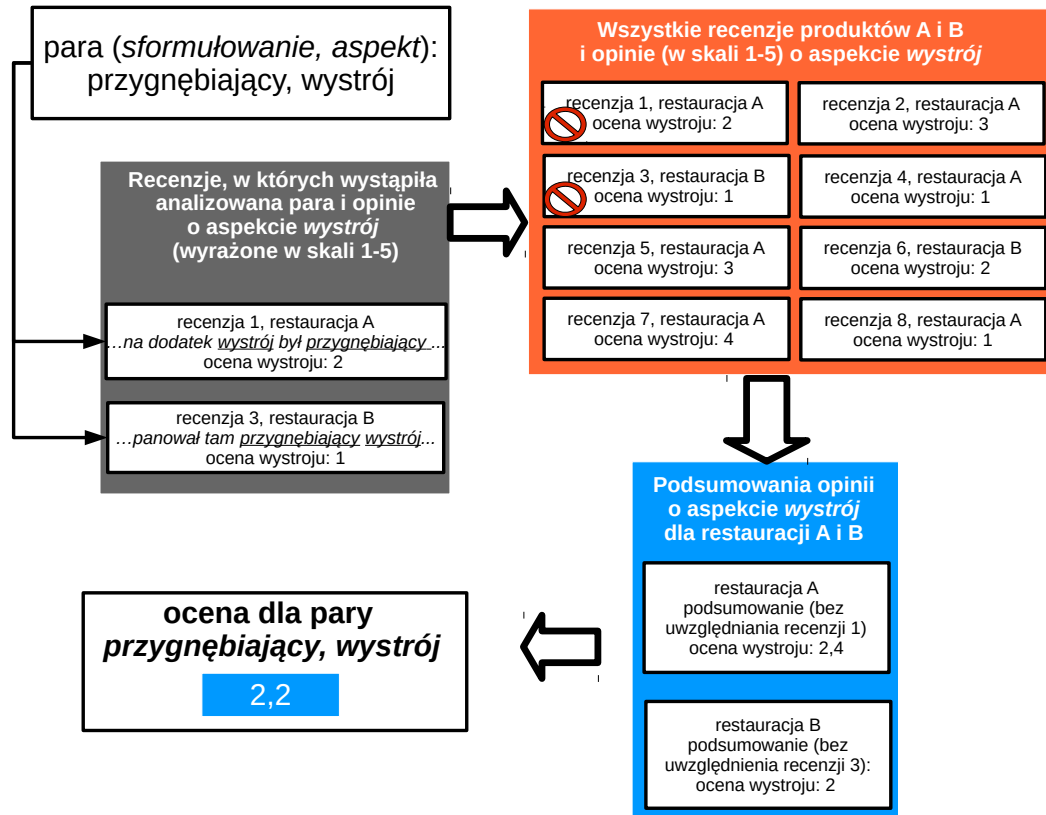
7.3.2 Przebieg eksperymentu i wykorzystane dane

Eksperyment przeprowadzany był na danych pobranych z portalu `gastronauci.pl`. Z portalu pobrano ponad 31000 recenzji restauracji działających na terenie Warszawy (każda recenzja była jednoznacznie przypisana do określonej restauracji). Każda recenzja składała się z trzech głównych elementów:

- tytułu (krótkiej wypowiedzi tekstowej),
- dłuższej wypowiedzi tekstowej,
- opinii wyrażonej w postaci oceny punktowej (w skali od 1 – najgorsza ocena do 5 – ocena najlepsza), przyznawanej następującym czterem aspektom restauracji:
 - jedzenie,
 - obsługa,
 - wystrój,
 - jakość/ceny.

Wypowiedzi tekstowe przetworzono w identyczny sposób, jak w przypadku recenzji telefonów komórkowych z portalu `cokupic.pl`. Na podstawie ręcznie przygotowanych list nazw aspektów i wypracowanych reguł ekstrakcji (analogicznych do tych opisanych w rozdziale 6), przeprowadzono ekstrakcję fraz, w których potencjalnie wyrażano opinie o aspektach recenzowanych dóbr. Na tej podstawie wygenerowano listę par (*sformułowanie, aspekt*), do których usiłowano przypisywać polarność.

Eksperyment miał analizować sytuację, gdy do par przypisywano polarność na podstawie podsumowań wielu opinii pochodzących ze źródła referencyjnego, jak to opisano w sekcji 4.3.1. Przypisywanie polarności nie odbywało się więc na podstawie ocen punktowych nadawanych przez autorów tych samych recenzji, w których dana para występowała, ale na podstawie podsumowań wielu opinii innych recenzentów (ze źródła referencyjnego) o restauracjach, w recenzjach których te pary się pojawiały. W eksperymencie symulowano tę sytuację w następujący sposób: dla każdej wyekstrahowanej frazy identyfikowano aspekt i restaurację, które podlegają ocenie. Następnie pobierano wszystkie opinie o danej restauracji wyrażone w postaci ocen punktowych za wyjątkiem oceny nadanej w tej recenzji, z której dana fraza została wyekstrahowana. W ten sposób uzyskiwano podsumowanie opinii wielu recenzentów o danym aspekcie, na które jednak nie wpływała ocena punktowa nadana przez autora recenzji, z której dana fraza została wyekstrahowana.



Rysunek 7.3: Proces przygotowywania danych dla potrzeb eksperymentu, za pomocą którego symulowano sytuację opisaną w opisie pierwszego wariantu proponowanej metody (patrz sekcja 4.3.1). Recenzje oznaczone znakiem zakazu nie są uwzględniane podczas generowania podsumowań, przez co symulowana jest sytuacja pobierania opinii ze źródła referencyjnego. Źródło: opracowanie własne

W wyniku uzyskiwano pojedynczą wartość, będącą średnią oceną danego aspektu dla usług, których dotyczyły recenzje, w których analizowana para się pojawiała, analogiczną do tego, jak opisano to w sekcji 4.3.1. Opisany proces generowania danych dla potrzeb eksperymentu zobrażowano na rysunku 7.3.

W rezultacie przeprowadzonej ekstrakcji pozyskano 651 par (*sformułowanie, aspekt*). Przeprowadzono ich ręczną adnotację, w której przypisywano do nich polarność, analogicznie jak to miało miejsce w eksperymencie opisanym w sekcji 7.1. Dwóch adnotatorów było zgodnych w 485 przypadkach i tym razem w eksperymentach wykorzystano jedynie te pary, dla których adnotatorzy byli ze sobą zgodni.

Do par przypisano uśrednione oceny z podsumowań, zgodnie z procedurą zaprezentowaną na rysunku 7.3 i opisaną w sekcji 4.3.1. Następnie analizowano, z jaką skutecznością możliwa jest identyfikacja polarności dla par na podstawie uzyskanych wartości. Jako podstawową cechę wykorzystywaną przy identyfikacji polarności, oprócz wartości średniej oceny, przyjęto liczbę wystąpień danej pary w korpusie. Jest to umotywowane tym, że im więcej razy dana para wystąpiła w korpusie, tym więcej podsumowań opinii o danym aspekcie możemy pozyskać i uzyskana końcowa ocena jest mniej podatna na przypadkowe wartości (co udowodniono w sekcji 5.4). Jeśli więc jakaś para pojawiła się 10 razy i średnia ocena danego aspektu w podsumowaniach wynosiła 2,4 to z większą pewnością możemy powiedzieć, że dana para prawdopodobnie ma negatywną polarność, niż gdyby taka średnia była wygenerowana na podstawie wyłącznie jednego podsumowania.

W związku z tym, przygotowano proste reguły, które miały przypisywać polarność do par na podstawie wspomnianych dwóch cech, czyli liczby wystąpień pary w korpusie i średniej oceny, nadanej danemu aspektowi w podsumowaniach opinii o produktach, w recenzjach których dana para występowała. Przygotowano osobne reguły dla identyfikacji polarności pozytywnej i negatywnej, które miały następującą formę: jeśli liczba wystąpień jest większa lub równa niż pewna graniczna wartość, a średnia ocena jest większa (lub mniejsza) niż kolejna, zdefiniowana dla niej wartość graniczna, do danej pary należy przypisać określoną polarność. Na przykład, reguła taka może przypisywać polarność negatywną do tych par, dla których liczba wystąpień w korpusie jest większa niż wartość graniczna trzech wystąpień oraz których średnia ocena była mniejsza niż wartość graniczna 2,75 (w skali do pięciu punktów). Inna reguła może przypisywać polarność pozytywną do tych par, które wystąpiły co najmniej 8 razy i średnia ocena wynosiła co najmniej 4.

Wykorzystując zdefiniowane reguły i przygotowane dane, przeprowadzono symulację poprawności identyfikacji polarności w zależności od przyjętych wartości granicznych. W trakcie symulacji, przyjmowano jedną z granicznych wartości jako ustaloną, natomiast drugą manipulowano i analizowano, jak zmienia się poprawność przypisywania polarności (poprawność ponownie ustalana była z wykorzystaniem miar precyzji i pełności). Eksperymenty osobno przeprowadzono dla polarności pozytywnej i negatywnej. Przypisywane polarności porównywano z polarnościami nadanymi przez adnotatorów.

7.3.3 Uzyskane wyniki i ich interpretacja

Uzyskane wyniki zaprezentowano na wykresach na rysunku 7.4 na stronie 155. Podczas interpretacji uzyskanych statystyk należy pamiętać, że proponowana metoda działa w scenariuszu, w którym istniejące metody nadzorowania odległego w ogóle nie byłyby w stanie przypisać polarności, ze względu na brak oceny przypisanej aspektowi przez autora danej recenzji. W związku z tym nawet jeśli proponowana metoda jest w stanie prawidłowo przypisać polarność jedynie w niewielkiej liczbie przypadków (ale przy wysokiej precyzji), to i tak stanowi to poprawę w stosunku do istniejących metod.

Ekspertyzy wykazały, że za pomocą proponowanej metody możliwe jest przeprowadzenie identyfikacji polarności z precyzją wahającą się pomiędzy 65% a nawet 100%, przy pełności nie przekraczającej 30%, w zależności od przypisywanej polarności, przyjętej oceny granicznej oraz minimalnej liczby wystąpień par, do których przypisywano polarność.

Przeprowadzone symulacje wyraźnie wskazały na wpływ zarówno przyjętej granicznej wartości oceny, jak również minimalnej liczby wystąpień par na precyzję i pełność identyfikacji polarności. Pierwszym wnioskiem płynącym z analizy zaprezentowanych wykresów jest to, że im bliżej wartości skrajnych ustalimy ocenę graniczną (czyli wartości 1 i wartości 5), tym wyższą precyzję jesteśmy w stanie uzyskać (przy jednoczesnym spadku pełności). Wniosek ten jest dość zrozumiały; w przypadku polarności pozytywnej, im bliżej oceny 5 jest średnia ocena, tym bardziej pozytywnie były postrzegane aspekty recenzowanych dóbr przez wielu recenzentów, a tym samym szansa na to, że dane sformułowanie wyraża wydźwięk pozytywny, jest wyższa.

Drugim wnioskiem z analizy wykresów jest to, że wraz ze wzrostem minimalnej liczby wystąpień wyraźnie następuje wzrost precyzji, przy jednoczesnym spadku pełności. Spadek pełności jest tu zjawiskiem naturalnym, ponieważ wraz ze wzrostem minimalnej liczby wystąpień odrzucono wszystkie pary, które występowały mniej razy niż zdefiniowana wartość. Natomiast wzrost precyzji wskazuje na fakt, że im więcej razy zaobserwowano daną parę i im więcej zebrano podsumowań opinii produktów, w recenzjach których dana para występowała, tym trafniej jesteśmy w stanie przypisywać polarność. Może być to spowodowane m.in. tym, że większa liczba podsumowań branych pod uwagę zmniejsza wpływ na końcową wartość pojedynczych podsumowań odstających od właściwego znaczenia danej pary (które zawsze mogą się zdarzać). Jest to zgodnie z wynikami uzyskanymi w eksperymencie 5.4 zaprezentowanymi na rysunku 5.3 na stronie 103.

Z przedstawionej zależności można wysnuć ważny wniosek dla wariantu metody opisanego w sekcji 4.3.1. W scenariuszu, system długotrwale analizuje teksty pojawiające się w zdefiniowanych

źródłach internetowych. Im dłużej system implementujący opisaną metodę będzie działał, tym więcej razy będzie mógł obserwować występowanie poszczególnych par i będzie miał większą liczbę podsumowań, na podstawie których będzie ustalał ostateczną średnią ocenę. Tym samym, system będzie dążył do coraz wyższej poprawności działania wraz z upływem czasu.

Na koniec, warto zwrócić uwagę na fakt, że praktycznie wszystkie przypadki niewłaściwej identyfikacji polarności (które pomniejszały precyzję rozwiązania) wiązały się z sytuacjami, gdy przypisano polarność pozytywną lub negatywną do par, które były oznaczane przez adnotatorów jako neutralne lub niewyrażające opinii. Innymi słowy, praktycznie nie zdarzała się sytuacja, w której przyznawano polarność pozytywną do pary, która faktycznie miała polarność negatywną lub odwrotnie. W związku z tym należy stwierdzić, że proponowana metoda działałaby znacznie lepiej dla leksykonu, z którego by usunięto pary o polarności neutralnej lub niewyrażające opinii. Podobnej obserwacji dokonano w pracy [Banea i in. 2008], gdzie stwierdzono, że prawidłowa identyfikacja sformułowań niewyrażających opinii mogłaby znacznie poprawić wyniki analizy wydźwięku (i identyfikacji polarności). Jednocześnie w tej samej pracy stwierdzono, że prawidłowe rozróżnianie fraz, w których jest wyrażona opinia od tych, w których opinia się nie pojawia, często może być zadaniem trudniejszym od samej analizy wydźwięku.

7.4 Podsumowanie

W rozdziale omówiono trzy przeprowadzone eksperymenty, które miały wykazać słuszność tezy przyjętej w pracy. Główne rezultaty zaprezentowano w tabelach 7.2, 7.3 i 7.6 oraz na rysunku 7.4. Pozwoliły one udowodnić, że:

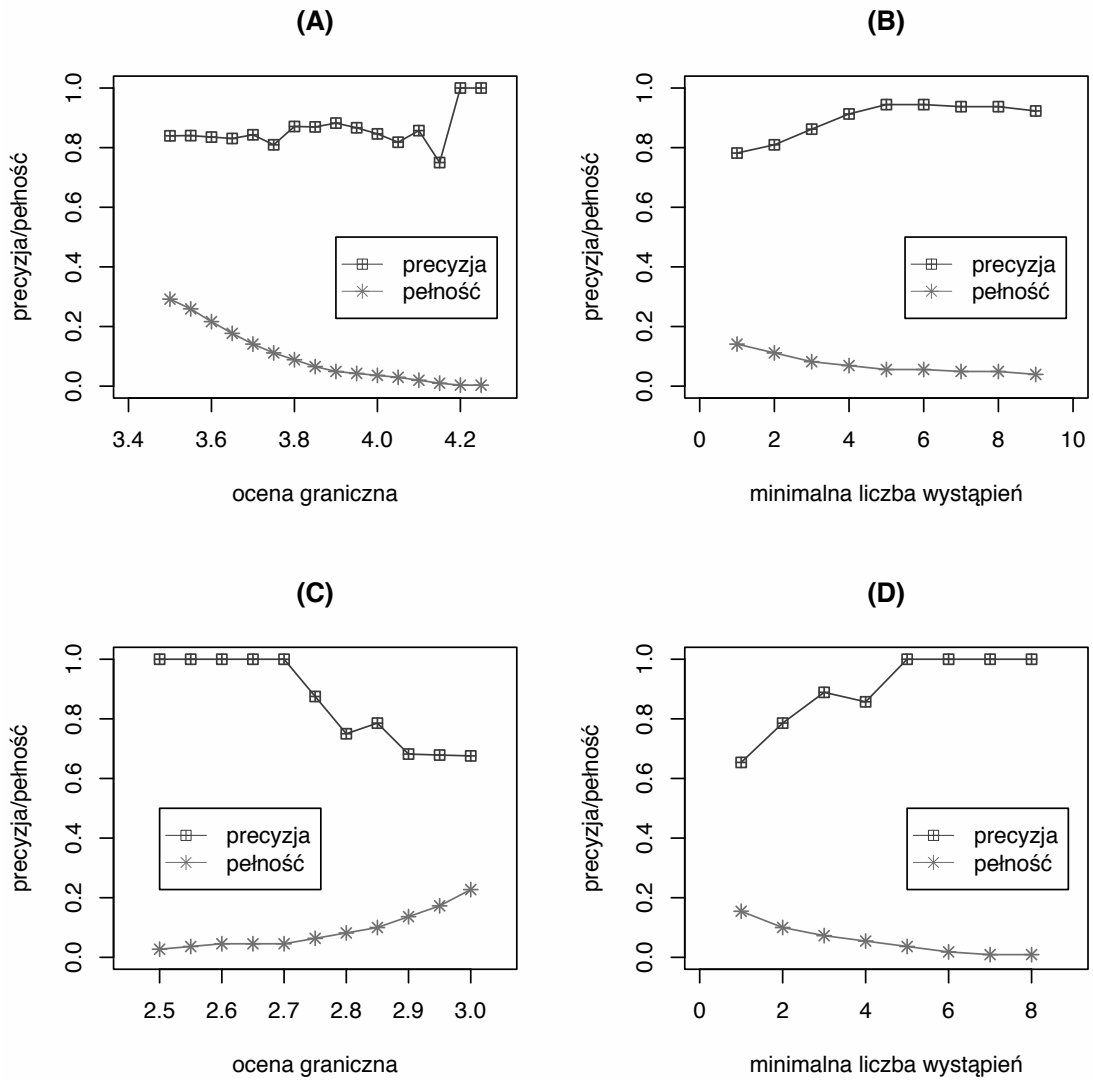
- w stosunku do istniejących rozwiązań, w scenariuszu, którego dotyczy drugi wariant proponowanej metody, opisany w sekcji 4.3.2, proponowana metoda pozwala na uzyskanie porównywalnych lub lepszych rezultatów identyfikacji polarności,
- również dla wariantu drugiego, leksykon uzyskany za pomocą proponowanej metody pozwala uzyskać lepsze rezultaty analizy wydźwięku od przetestowanych dwóch metod bazowych,
- proponowana metoda pozwala identyfikować polarność sformułowań z wypowiedzi tekstowych również w sytuacjach, gdy dana wypowiedź nie posiada przypisanej opinii wyrażonej w strukturyzowany sposób; jest to możliwe dzięki wykorzystaniu podsumowań wielu przetwarzalnych opinii o recenzowanym produkcie lub usłudze, pobranych z innych źródeł,

- leksykon uzyskany z wykorzystaniem zaproponowanej metody pozwala na uzyskanie lepszych rezultatów analizy wydźwięku niż przyjęte rozwiązania bazowe.

Przeprowadzone eksperymenty pozwoliły więc na osiągnięcie drugiego i trzeciego celu szczegółowego i udowodniły tezę przyjętą w pracy.

polarność	statystyka	bazowa 1	bazowa 2	prezentowana w rozprawie
pozytywna	precyzja	75,46%	86,55%	80,75%
	pełność	98,19%	62,05%	90,96%
	F_1	85,34%	72,28%	85,55%
	$F_{0,5}$	79,13%	80,22%	82,6%
negatywna	precyzja	72,22%	73,53%	70%
	pełność	43,33%	41,67%	46,67%
	F_1	54,17%	53,19%	56%
	$F_{0,5}$	63,73%	63,78%	63,64%
neutralna / brak opinii	precyzja	25,0%	25,0%	34,48%
	pełność	1,25%	62,5%	25,0%
	F_1	2,38%	35,71%	28,99%
	$F_{0,5}$	5,21%	28,41%	32,05%
wszystkie	$\overline{F_1}$	67,32	63,66%	72,21%
	$\overline{F_{1PN}}$	80,56%	69,36%	81,03%
	$\overline{F_{0,5}}$	64,64%	69,34%	71,62%
	$\overline{F_{0,5PN}}$	76,77%	77,7%	79,7%

Tabela 7.6: Statystyki precyzji, pełności i miary F dla poszczególnych polarności, obrazujące poprawność analizy wydźwięku przeprowadzonej z wykorzystaniem leksykonów będących wynikiem trzech porównywanych metod. Źródło: opracowanie własne



Rysunek 7.4: Precyzja i pełność proponowanej metody dla polarności pozytywnej ((A) i (B)) i negatywnej ((C) i (D)) w zależności od przyjętej wartości oceny granicznej (przy minimalnej liczbie wystąpień na poziomie 3) i w zależności od minimalnej liczby wystąpień par poddawanych klasyfikacji (przy granicznej wartości oceny na poziomie 3,75 dla polarności pozytywnej i 2,85 dla polarności negatywnej). Źródło: opracowanie własne

Rozdział 8

Podsumowanie

Przedmiotem prezentowanej rozprawy była analiza wydźwięku, tj. zadanie mające na celu automatyczne przetwarzanie opinii zawartych w wypowiedziach tekstowych. Jak wykazano w pracy, zagadnienie to może pełnić istotną rolę w funkcjonowaniu podmiotów działających na współczesnym rynku. Praca skupiała się na ekstrakcji leksykonów sformułowań służących do wyrażania opinii, a dokładniej na identyfikacji polarności takich sformułowań. Takie leksykony są wykorzystywane przez wiele podejść do analizy wydźwięku.

Analiza wydźwięku od wielu lat pozostaje w obszarze zainteresowania informatyki ekonomicznej. Popularnym nurtem badań w tej dziedzinie jest nadzorowanie odległe, tj. jednoczesne analizowanie recenzji sformułowanych w postaci wypowiedzi tekstowych oraz w dodatkowej, łatwo przetwarzalnej postaci. Aby zastosować opisane podejście wymagane jest, aby opinie wyrażone na wspomniane dwa sposoby były wyrażone w ramach jednej recenzji, tj. przez tego samego recenzenta w tym samym momencie. Wówczas, opinie wyrażone w przetwarzalny sposób mogą być traktowane jako wskazówki przy ustalaniu polarności sformułowań używanych w wypowiedziach tekstowych.

W prezentowanej pracy usiłowano wprowadzić usprawnienie do opisanego podejścia.

8.1 Wkład pracy

Głównym celem pracy było opracowanie metody, wykorzystującej informacje zawarte w wielu częściowo strukturyzowanych recenzjach, w celu poprawy skuteczności istniejących metod identyfikacji polarności sformułowań służących do wyrażania opinii, dla potrzeb oceny użyteczności produktów i usług. Taki cel pracy został zdefiniowany na podstawie przeprowadzonej analizy

literatury z zakresu wpływu recenzji konsumenckich na zachowania podmiotów uczestniczących w wymianie rynkowej i automatycznego przetwarzania opinii zawartych w wypowiedziach tekstowych (drażnienia opinii), a także analizy portali umożliwiających publikowanie recenzji konsumenckich.

Główny cel pracy miał zostać osiągnięty poprzez realizację trzech celów szczegółowych. Te cele to:

1. Analiza możliwości wykorzystania podsumowań wielu opinii o produktach i usługach dla potrzeb ustalania polarności sformułowań wykorzystywanych przez recenzentów do wyrażania opinii,
2. Opracowanie metody identyfikacji polarności sformułowań wyrażających opinie o aspektach produktów i usług na podstawie informacji pozyskanych z podsumowań wielu recenzji konsumenckich, oceniających te produkty i usługi,
3. Opracowanie metody analizy wydźwięku wykorzystującej informacje zawarte w podsumowaniach wielu opinii o produktach i usługach.

Każdy z tych celów został osiągnięty poprzez wypracowanie odpowiednich artefaktów oraz przeprowadzenie ich ewaluacji:

- Cel szczegółowy 1 został osiągnięty na drodze analizy przykładowego korpusu recenzji konsumenckich. Za pomocą przeprowadzonych testów statystycznych wykazano, że:
 - istnieją sytuacje, w których nie ma wystarczających danych do przeprowadzenia identyfikacji polarności w ramach tradycyjnego podejścia do nadzorowania odległego,
 - opinie wielu recenzentów na temat pojedynczego produktu (lub pojedynczej usługi) są ze sobą co do zasady spójne, zarówno, gdy wyrażane są tylko w jednym formacie, jak również wówczas, gdy są wyrażane za pomocą różnych sposobów wyrażania opinii. Dzięki temu możliwe jest przeprowadzanie wnioskowań o produktach i usługach na podstawie takich podsumowań. W kontekście prezentowanej pracy oznacza to, że znając podsumowanie opinii wielu recenzentów o danym produkcie, można z pewnym prawdopodobieństwem przewidzieć, jaką opinię może o tym produkcie wyrazić inny recenzent, a tym samym jaka jest polarność sformułowania użytego przez niego do wyrażenia opinii.

- Cel szczegółowy 2 został zrealizowany poprzez wypracowanie modeli, metod oraz implementację prototypów, za pomocą których przeprowadzono identyfikację polarności na podstawie podsumowań opinii wielu recenzentów o produktach i usługach. Szczegółowo opisano wypracowane metody, a także przygotowano implementację potoku przetwarzania umożliwiającego przeprowadzenie zaproponowanych wnioskowań. Wypracowane artefakty zostały poddane ewaluacji, w której porównano podejście przedstawione w rozprawie z dwoma wybranymi rozwiązaniami (tzw. metodami bazowymi), reprezentującymi dotychczasowe podejścia do identyfikacji polarności z wykorzystaniem nadzorowania odległego. Ewaluacja wykazała, że proponowane w pracy podejście pozwala na uzyskanie lepszych rezultatów identyfikacji polarności.
- Cel szczegółowy 3 osiągnięto poprzez przeprowadzenie analizy wydźwięku z wykorzystaniem leksykonów uzyskanych za pomocą trzech metod wspomnianych w poprzednim punkcie i wykazanie, że wykorzystanie leksykonu wypracowanego za pomocą proponowanego podejścia pozwoliło na uzyskanie najlepszych rezultatów.

Osiągając przedstawione cele, wykazano tezę pracy:

Wykorzystanie podsumowania opinii recenzentów o pewnym produkcie lub usłudze, skonstruowanego na podstawie wielu częściowo strukturyzowanych recenzji tego produktu (lub usługi), poprawi skuteczność budowania leksykonu sformułowań wyrażających opinie o różnych jego aspektach, a tym samym pozwoli na opracowanie metody oceny użyteczności produktów i usług przez konsumentów.

W trakcie prac służących osiągnięciu celu głównego pracy i wykazaniu jej tezy osiągnięto również poboczne rezultaty, które również stanowią wkład do analizowanej dziedziny:

- W trakcie prac nad osiągnięciem celu szczegółowego 1 wypracowano miarę odległości pomiędzy podsumowaniami wielu opinii o produktach i usługach. Wyniki eksperymentów potwierdziły, że wykorzystując tę metrykę do pomiaru odległości możliwe jest wykazanie, że odległości między podsumowaniami opinii o tym samym produkcie są statystycznie mniejsze niż pomiędzy podsumowaniami produktów różnych. Pozwala to oczekiwać, że miara w przypadku obliczania odległości pomiędzy różnymi produktami, które jednak są podobnie postrzegane przez recenzentów, również będzie zwracała mniejsze odległości niż w przypadku obliczania odległości pomiędzy produktami postrzeganymi inaczej. Wniosek taki można

wyciągnąć wówczas, gdy dwa podsumowania tego samego produktu, wygenerowane na podstawie różnych jego recenzji, potraktuje się jako szczególny przypadek sytuacji, gdy dysponujemy dwoma podsumowaniami różnych produktów postrzeganych przez recenzentów w podobny sposób. Wypracowana miara może zostać wykorzystana do przeprowadzania dalszych wnioskowań na generowanych podsumowaniach wielu opinii o produktach i usługach.

- Wypracowano trójfazową metodę ekstrakcji z tekstu sformułowań służących do wyrażania opinii wraz z aspektami, których te sformułowania dotyczą; zaproponowana trzecia faza ekstrakcji, polegająca na dodatkowym przefiltrowaniu wyekstrahowanych sformułowań z wykorzystaniem uczenia maszynowego, pozwala na podniesienie precyzji ekstrakcji, a tym samym na generowanie poprawniejszych leksykonów.

8.2 Dalsze badania

Dalsze badania mogą toczyć się wokół wykorzystania kolejnych formatów wyrażania opinii, przykładowo list zalet i wad formułowanych za pomocą języka naturalnego czy wyłącznie działając z wykorzystaniem wypowiedzi tekstowych, tak jak opisano to w rozdziale 4, przy okazji opisywania trzeciego wariantu proponowanej metody. Zaproponowano w nim, aby generować podsumowania opinii wyrażanych w postaci wypowiedzi tekstowych i takie podsumowania wykorzystywać w procesie identyfikacji polarności. Scenariusz taki jest znacznie trudniejszy do realizacji od sytuacji opisanych w pozostałych wariantach i wymaga możliwości przeprowadzenia analizy wydźwięku na wysokim poziomie poprawności na bardzo dużych kolekcjach recenzji. Jego eksperymentalne przeanalizowanie byłoby interesującym zadaniem badawczym.

Kolejnym kierunkiem dalszych badań może być próba integracji zaproponowanego podejścia z innymi rozwiązaniami z zakresu identyfikacji polarności, korzystającymi z innych przesłanek do przeprowadzenia tej operacji. Łączenie wielu takich przesłanek może pozwolić na połączenie korzyści płynących z różnych podejść (przykładem pracy mającej taki właśnie cel jest [Lu i in. 2011]).

Wyniki prezentowanej pracy mogą zostać wykorzystane do generowania podsumowań opinii o poszczególnych aspektach recenzowanych bytów. Dla wielu podmiotów, które mogą wykorzystać takie podsumowania, istotną funkcją mogłaby być możliwość łatwego przeprowadzania pogłębionej analizy opinii, które zostały wykorzystane do wygenerowania prezentowanych podsumowań. Podobne funkcje pogłębionej analizy określonych danych są często dostępne w systemach anali-

tycznych i określane są mianem *drill down*. Przykładowo, użytkownicy, którym zaprezentowano raport z przeprowadzonych analiz, poprzez zaznaczenie wybranego elementu raportu (np. określonej serii danych na wykresie), mogliby mieć możliwość analizy fragmentów zdań, które zidentyfikowano jako wyrażające opinię o określonym produkcie lub usłudze, przykładowo w celu poznania przyczyn wyrażenia określonej (pozytywnej lub negatywnej) opinii. Z opisaną funkcją wiąże się kolejne interesujące zadanie badawcze, polegające na automatycznym podsumowaniu przyczyn (uzasadnień), dla których recenzenci wyrażali określone opinie o produktach, usługach i ich aspektach. Wynikiem metod wypracowanych w tym zakresie mogłoby być prezentowanie graficznych podsumowań takich przyczyn, co mogłoby ułatwić podmiotom podejmowanie właściwych decyzji na podstawie przeprowadzanych analiz.

W trakcie prowadzonych prac nad analizą spójności postrzegania określonych produktów i usług przez wielu konsumentów, wypracowano miarę odległości pomiędzy podsumowaniami wielu opinii wyrażonych o konkretnych produktach i usługach. Wykazano, że korzystając z zaproponowanej miary, podsumowania wielu różnych opinii dotyczących tych samych dóbr są do siebie podobne. Otrzymane rezultaty eksperymentów pozwalają oczekiwać, że takie podobieństwo można badać również pomiędzy podsumowaniami różnych produktów i usług. Interesującym wyzwaniem byłoby wyszukiwanie dóbr podobnie postrzeganych przez recenzentów, przykładowo w celu rekomendacji konsumentom podobnych produktów po niższej cenie. Kolejnym zastosowaniem takiej analizy mogłoby być wyszukiwanie przez pośredników dóbr podobnych do określonego produktu lub usługi, oferowanych przez innego producenta, co mogłoby pomóc tym pośrednikom w odpowiednim kształtowaniu relacji z dostawcami.

Zagadnienie to można poprowadzić jeszcze dalej. Przykładowo, interesującym zadaniem mogłoby być wyszukiwanie produktów, które są w wielu aspektach do siebie podobne, ale w określonym aspekcie jeden jest lepszy od drugiego. Eksperymenty przeprowadzone w prezentowanej pracy pozwalają oczekiwać, że pozyskiwane w ten sposób informacje o relacjach pomiędzy produktami i usługami mogłyby odwzorowywać faktyczne podobieństwa pomiędzy nimi, a wykorzystanie tych informacji mogłoby przynieść korzyść określonym podmiotom podczas podejmowania decyzji. Kolejnym powiązaniem zagadnieniem z kierunkiem opisanym w poprzednim akapicie może być przeprowadzanie analiz na jeszcze wyższym poziomie ogólności, mających na celu analizę postrzegania całych koszyków produktów i usług. Analiza taka może przykładowo pozwolić pośrednikom na analizę sposobu postrzegania całej oferty określonego producenta.

Interesującym wykorzystaniem otrzymanych rezultatów badań i rozwinięciem proponowanego

podejścia może być także wykorzystanie analizy podsumowań wielu opinii o produktach i usługach do celów identyfikacji aspektów i przypisywania różnych nazw do właściwych aspektów. Pojedynczy aspekt może mieć wiele nazw, które często są niepowiązanymi ze sobą wyrazami. Jak wykazano w prezentowanej pracy, podsumowania opinii wielu recenzentów o produkcie są ze sobą spójne. Interesującym wykorzystaniem tej obserwacji mogłoby być wygenerowanie podsumowań opinii przypisywanych do poszczególnych nazw aspektów, o których nie wiadomo, czy odnoszą się do tego samego aspektu. Jeśli takie podsumowania byłyby ze sobą spójne w wielu przypadkach, to można z dużym prawdopodobieństwem traktować je jako różne nazwy tego samego aspektu.

Bibliografia

- Abramowicz, W., 2008, *Filtrowanie informacji*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań.
- Andreevskaia, A. i Bergler, S., 2006, *Mining WordNet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses*, w: *Proceedings of EACL*, tom 6, s. 209–216.
- Arora, R. i Srinivasa, S., 2014, *A faceted characterization of the opinion mining landscape.*, w: *COMSNETS*, s. 1–6.
- Baccianella, S., Esuli, A. i Sebastiani, F., 2010, *Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining*, w: *Seventh conference on International Language Resources and Evaluation, Malta. Retrieved May*, tom 25, s. 2010.
- Balabanović, M. i Shoham, Y., 1997, *Fab: content-based, collaborative recommendation*, *Commun. ACM*, tom 40, nr 3, s. 66–72, ISSN 0001-0782.
- Balahur, A. i Turchi, M., 2012, *Multilingual Sentiment Analysis Using Machine Translation?*, w: *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA '12, Association for Computational Linguistics, Stroudsburg, PA, USA, s. 52–60, URL <http://dl.acm.org/citation.cfm?id=2392963.2392976>.
- Banea, C., Mihalcea, R., Wiebe, J. i Hassan, S., 2008, *Multilingual subjectivity analysis using machine translation*, w: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, s. 127–135.
- Baran, T., 2000, *Jak satysfakcjonować klienta*, *Marketing w Praktyce*, tom 1, nr 1, s. 11–13.
- Barton, B., 2006, *Ratings, reviews & ROI: How leading retailers use customer word of mouth in marketing and merchandising*, *Journal of Interactive Advertising*, tom 7, nr 1, s. 1–7.

- Bernatowicz, A. i Małyszko, J., 2014, *Recenzje konsumenckie w Internecie. Społeczny kontekst publikowania opinii i analiza spójności różnych sposobów ich wyrażania*, w: *Matematyka i informatyka na usługach ekonomii. Teoria i zastosowania*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, ISBN 978-83-7417-805-1, s. 158–168–7:8.
- Bickart, B. i Schindler, R., 2001, *Internet forums as influential sources of consumer information*, *Journal of interactive marketing*, tom 15, nr 3, s. 31–40.
- Bilenko, M. i Mooney, R.J., 2003, *Adaptive Duplicate Detection Using Learnable String Similarity Measures*, w: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, ACM, New York, NY, USA, ISBN 1-58113-737-0, s. 39–48, URL <http://doi.acm.org/10.1145/956750.956759>.
- Bonev, B., Ramírez-Sánchez, G. i Rojas, S.O., 2012, *Opinum: Statistical Sentiment Analysis for Opinion Classification*, w: *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA '12, Association for Computational Linguistics, Stroudsburg, PA, USA, s. 29–37, URL <http://dl.acm.org/citation.cfm?id=2392963.2392973>.
- Broß, D.I.J., 2013, *Aspect-Oriented Sentiment Analysis of Customer Reviews Using Distant Supervision Techniques*, Rozprawa doktorska, Berlin, Freie Universität Berlin, Diss., 2013.
- Buczynski, A. i Wawer, A., 2008a, *Automated classification of product review sentiments in Polish*, w: *Intelligent Information Systems*, tom XVI.
- Buczynski, A. i Wawer, A., 2008b, *Shallow parsing in sentiment analysis of product reviews*, w: *Proceedings of the Sixth International Conference on Language Resources and Evaluation*.
- Carenini, G., Ng, R.T. i Zwart, E., 2005, *Extracting knowledge from evaluative text*, w: *Proceedings of the 3rd international conference on Knowledge capture*, K-CAP '05, ACM, New York, NY, USA, ISBN 1-59593-163-5, s. 11–18.
- Chakravarthy, S. i Prasad, G.V., 2011, *The Impact of Opinion Leader on Consumer Decision Making Process*, *International Journal of Management & Business Studies*, tom 1, nr 3, s. 61–64.

- Chan, K.T. i King, I., 2009, *Let's Tango — Finding the Right Couple for Feature-Opinion Association in Sentiment Analysis*, w: *Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, PAKDD '09, Springer-Verlag, Berlin, Heidelberg, ISBN 978-3-642-01306-5, s. 741–748, URL http://dx.doi.org/10.1007/978-3-642-01307-2_75.
- Chaovalit, P. i Zhou, L., 2005, *Movie review mining: A comparison between supervised and unsupervised classification approaches*, w: *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*, IEEE, s. 112c–112c.
- Chen, L., Qi, L. i Wang, F., 2012, *Comparison of feature-level learning methods for mining online consumer reviews*, *Expert Syst. Appl.*, tom 39, nr 10, s. 9588–9601, ISSN 0957-4174.
- Chen, S. i Goodman, J., 1996, *An empirical study of smoothing techniques for language modeling*, w: *Proceedings of the 34th annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics, s. 310–318.
- Chen, Y. i Xie, J., 2005, *Third-party product review and firm marketing strategy*, *Marketing Science*, s. 218–240.
- Chevalier, J.A. i Mayzlin, D., 2003, *The effect of word of mouth on sales: Online book reviews*, Rap. tech., National Bureau of Economic Research.
- Choi, Y. i Cardie, C., 2009, *Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification*, w: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*, Association for Computational Linguistics, s. 590–598.
- Cohen, W.W., Ravikumar, P.D., Fienberg, S.E. i in., 2003, *A Comparison of String Distance Metrics for Name-Matching Tasks.*, w: *IWeb*, tom 2003, s. 73–78.
- comScore, 2007, *Online Consumer-Generated Reviews Have Significant Impact on Offline Purchase Behavior*, http://www.comscore.com/Press_Events/Press_Releases/2007/11/Online_Consumer_Reviews_Impact_Offline_Purchasing_Behavior.
- Cui, H., Mittal, V. i Datar, M., 2006, *Comparative experiments on sentiment classification for online product reviews*, w: *Proceedings of the National Conference on Artificial Intelligence*, nr 2, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, s. 1265.

- Cunningham, H., Maynard, D. i Tablan, V., 1999, *JAPE: a Java Annotation Patterns Engine. Research memo.*
- Dalal, M.K. i Zaveri, M.A., 2014, *Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges*, Appl. Comp. Intell. Soft Comput., tom 2014, s. 2:2–2:2, ISSN 1687-9724, URL <http://dx.doi.org/10.1155/2014/735942>.
- Dave, K., Lawrence, S. i Pennock, D.M., 2003, *Mining the peanut gallery: opinion extraction and semantic classification of product reviews*, w: *Proceedings of the 12th international conference on World Wide Web*, WWW '03, ACM, New York, NY, USA, ISBN 1-58113-680-3, s. 519–528.
- Dellarocas, C., 2006, *Strategic manipulation of internet opinion forums: Implications for consumers and firms*, Management Science, s. 1577–1593.
- Dellarocas, C., Zhang, X. i Awad, N., 2007, *Exploring the value of online product reviews in forecasting sales: The case of motion pictures*, Journal of Interactive Marketing, tom 21, nr 4, s. 23–45.
- Denecke, K., 2008, *Using SentiWordNet for multilingual sentiment analysis*, w: *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on*, IEEE, s. 507–512.
- Ding, X., Liu, B. i Yu, P.S., 2008, *A holistic lexicon-based approach to opinion mining*, w: *Proceedings of the 2008 International Conference on Web Search and Data Mining*, WSDM '08, ACM, New York, NY, USA, ISBN 978-1-59593-927-2, s. 231–240.
- Dzikowski, J., Filipowska, A., Łazaruk, S. i Abramowicz, W., 2011, *Towards the Semantic Web's application for preparation of reviews—requirements and architecture for the needs of incentive-based semantic content creation*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, , nr 206, s. 11.
- Esuli, A. i Sebastiani, F., 2005, *Determining the semantic orientation of terms through gloss classification*, w: *Proceedings of the 14th ACM international conference on Information and knowledge management*, CIKM '05, ACM, New York, NY, USA, ISBN 1-59593-140-6, s. 617–624, URL <http://doi.acm.org/10.1145/1099554.1099713>.
- Frank, E., Paynter, G., Witten, I., Gutwin, C. i Nevill-Manning, C., 1999, *Domain-specific keyphrase extraction*, Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence.

- Gamon, M., 2004, *Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis*, w: *Proceedings of the 20th international conference on Computational Linguistics*, Association for Computational Linguistics, s. 841.
- Gamon, M., Aue, A., Corston-Oliver, S. i Ringger, E., 2005, *Pulse: Mining customer opinions from free text*, *Advances in Intelligent Data Analysis VI*, s. 741–741.
- Gaul, W. i Schmidt-Thieme, L., 2001, *Recommender Systems Based on Navigation Path Features*, w: *Proceedings of WEBKDD Workshop at the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Ghose, A. i Ipeirotis, P.G., 2007, *Designing novel review ranking systems: predicting the usefulness and impact of reviews*, w: *Proceedings of the ninth international conference on Electronic commerce*, ICEC '07, ACM, New York, NY, USA, ISBN 978-1-59593-700-1, s. 303–310.
- Giménez, H., 2010, *Machine Learning - Who's the Boss?*
- Godbole, N., Srinivasaiah, M. i Skiena, S., 2007, *Large-scale sentiment analysis for news and blogs*, w: *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, s. 219–222.
- Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., Libai, B., Sen, S., Shi, M. i Verleghe, P., 2005, *The firm's management of social interactions*, *Marketing Letters*, tom 16, nr 3, s. 415–428.
- Grefenstette, G., Qu, Y., Evans, D. i Shanahan, J., 2006, *Validating the Coverage of Lexical Resources for Affect Analysis and Automatically Classifying New Words along Semantic Axes*, w: J. Shanahan, Y. Qu i J. Wiebe (red.), *Computing Attitude and Affect in Text: Theory and Applications*, tom 20 z serii *The Information Retrieval Series*, Springer Netherlands, ISBN 978-1-4020-4102-0, s. 93–107.
- Gregor, S. i Hevner, A.R., 2013, *Positioning and presenting design science research for maximum impact*, *MIS Quarterly*, tom 37, nr 2, s. 337–356.
- Gretzel, U. i Yoo, K., 2008, *Use and impact of online travel reviews*, *Information and communication technologies in tourism 2008*, s. 35–46.

- Harmon, A., 2004, *Amazon Glitch Unmasks War Of Reviewers*, The New York Times, <http://www.nytimes.com/2004/02/14/us/amazon-glitch-unmasks-war-of-reviewers.html?pagewanted=all&src=pm>, dostęp 10.09.2012.
- Hatzivassiloglou, V. i McKeown, K.R., 1997, *Predicting the semantic orientation of adjectives*, w: *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, EACL '97, Association for Computational Linguistics, Stroudsburg, PA, USA, s. 174–181, URL <http://dx.doi.org/10.3115/979617.979640>.
- Haulrich, M., 2009, *Comparing Classifiers for (a special case of) Transition-Based Parsing*, stp.lingfil.uu.se/~nivre/gslt/haulrich.pdf.
- Hennig-Thurau, T., Gwinner, K., Walsh, G. i Gremler, D., 2004, *Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?*, *Journal of interactive marketing*, tom 18, nr 1, s. 38–52.
- Hevner, A., March, S.T., Park, J. i Ram, S., 2004, *Design science in information systems research*, *MIS quarterly*, tom 28, nr 1, s. 75–105.
- Hevner, A.R., 2007, *A three cycle view of design science research*, *Scandinavian journal of information systems*, tom 19, nr 2, s. 4.
- Higashinaka, R., Prasad, R. i Walker, M.A., 2006, *Learning to generate naturalistic utterances using reviews in spoken dialogue systems*, w: *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, ACL-44, Association for Computational Linguistics, Stroudsburg, PA, USA, s. 265–272.
- Hill, N. i Alexander, J., 2003, *Pomiar satysfakcji i lojalności klientów*, Oficyna Ekonomiczna, Kraków.
- Ho, J. i Tang, R., 2001, *Towards an optimal resolution to information overload: an infomediary approach*, w: *GROUP '01: Proceedings of the 2001 International ACM SIGGROUP Conference on Supporting Group Work*, ACM, New York, NY, USA, ISBN 1-58113-294-8, s. 91–96.
- Hogenboom, A., Heerschop, B., Frasincar, F., Kaymak, U. i de Jong, F., 2014, *Multi-lingual support for lexicon-based sentiment analysis guided by semantics*, *Decision Support Systems*, tom 62, s. 43–53.

- Hu, M. i Liu, B., 2004a, *Mining and summarizing customer reviews*, w: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, s. 168–177.
- Hu, M. i Liu, B., 2004b, *Mining opinion features in customer reviews*, w: *Proceedings of the National Conference on Artificial Intelligence*, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, s. 755–760.
- Hu, N., Liu, L. i Zhang, J., 2008, *Do online reviews affect product sales? The role of reviewer characteristics and temporal effects*, *Information Technology and Management*, tom 9, nr 3, s. 201–214.
- Hu, N., Pavlou, P. i Zhang, J., 2006, *Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of Online word-of-mouth communication*, w: *Proceedings of the 7th ACM conference on Electronic commerce*, ACM, s. 324–330.
- Hu, N., Pavlou, P.A. i Zhang, J., 2007, *Why do online product reviews have a J-shaped distribution? Overcoming biases in online word-of-mouth communication*, *Marketing Science*, tom 198, s. 7.
- Hu, N., Zhang, J. i Pavlou, P.A., 2009, *Overcoming the J-shaped distribution of product reviews*, *Commun. ACM*, tom 52, nr 10, s. 144–147, ISSN 0001-0782, URL <http://doi.acm.org/10.1145/1562764.1562800>.
- Hu, Y., Lu, R., Li, X., Chen, Y. i Duan, J., 2007, *A Language Modeling Approach to Sentiment Analysis*, w: Y. Shi, G. Albada, J. Dongarra i P. Sloot (red.), *Computational Science – ICCS 2007*, tom 4488 z serii *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, ISBN 978-3-540-72585-5, s. 1186–1193.
- Jindal, N. i Liu, B., 2007, *Review spam detection*, w: *Proceedings of the 16th international conference on World Wide Web, WWW '07*, ACM, New York, NY, USA, ISBN 978-1-59593-654-7, s. 1189–1190, URL <http://doi.acm.org/10.1145/1242572.1242759>.
- Jędrzejewski, K. i Morzy, M., 2011, *Opinion Mining and Social Networks: A Promising Match*, w: *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, IEEE, s. 599–604.

- Joachims, T., 1998, *Text categorization with support vector machines: Learning with many relevant features*, Machine learning: ECML-98, s. 137–142.
- Joachims, T., 1999, *Making large scale SVM learning practical*, Advances in kernel methods: support vector learning.
- Kaczmarek, M., Filipowska, A., Dzikowski, J., Łazaruk, S. i Abramowicz, W., 2012, *Taste It! Try It!—A Semantic Web Mobile Review Application*, w: *eKNOW 2012, The Fourth International Conference on Information, Process, and Knowledge Management*, s. 122–127.
- Kaji, N. i Kitsuregawa, M., 2007, *Building Lexicon for Sentiment Analysis from Massive Collection of HTML Documents.*, w: *EMNLP-CoNLL*, s. 1075–1083.
- Kamps, J., Marx, M., Mokken, R. i De Rijke, M., 2004, *Using wordnet to measure semantic orientations of adjectives*, Proceedings of LREC.
- Kanayama, H. i Nasukawa, T., 2006, *Fully automatic lexicon expansion for domain-oriented sentiment analysis*, w: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP '06*, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 1-932432-73-6, s. 355–363.
- Kim, S.M. i Hovy, E., 2004, *Determining the sentiment of opinions*, w: *Proceedings of the 20th international conference on Computational Linguistics, COLING '04*, Association for Computational Linguistics, Stroudsburg, PA, USA.
- Kowalska, K., Cai, D. i Wade, S., 2012, *Sentiment Analysis of Polish Texts*, International Journal of Computer and Communication Engineering, tom 1, nr 1, s. 39–42.
- Łazaruk, S., Dzikowski, J., Kaczmarek, M. i Abramowicz, W., 2012, *Semantic Web Recommendation Application*, Proceedings of the Federated Conference on Computer Science and Information Systems.
- Li, M., Qi, Z., Lu, K. i Guo, C., 2014, *A textual polarity analysis based on reviewer identity disclosure and product sales*, w: *Computer Science & Education (ICCSE), 2014 9th International Conference on*, IEEE, s. 303–308.
- Linden, G., Smith, B. i York, J., 2003, *Amazon.com Recommendations: Item-to-Item Collaborative Filtering*, IEEE Internet Computing, tom 7, nr 1, s. 76–80, ISSN 1089-7801.

- Liu, B., 2011, *Opinion Mining and Sentiment Analysis*, w: *Web Data Mining*, Data-Centric Systems and Applications, Springer Berlin Heidelberg, ISBN 978-3-642-19460-3, s. 459–526.
- Liu, B., Hu, M. i Cheng, J., 2005, *Opinion observer: analyzing and comparing opinions on the Web*, w: *Proceedings of the 14th international conference on World Wide Web*, WWW '05, ACM, New York, NY, USA, ISBN 1-59593-046-9, s. 342–351, URL <http://doi.acm.org/10.1145/1060745.1060797>.
- Liu, L., Liu, H., Wang, H., Song, W. i Lu, J., 2014, *The study of sentiment analysis on Chinese comment texts*, w: *Software Engineering and Service Science (ICSESS), 2014 5th IEEE International Conference on*, IEEE, s. 397–401.
- Lu, Y., Castellanos, M., Dayal, U. i Zhai, C., 2011, *Automatic construction of a context-aware sentiment lexicon: an optimization approach*, w: *Proceedings of the 20th international conference on World wide web*, WWW '11, ACM, New York, NY, USA, ISBN 978-1-4503-0632-4, s. 347–356.
- Lu, Y., Zhai, C. i Sundaresan, N., 2009, *Rated aspect summarization of short comments*, w: *Proceedings of the 18th international conference on World wide web*, WWW '09, ACM, New York, NY, USA, ISBN 978-1-60558-487-4, s. 131–140.
- Lula, P. i Wójcik, K., 2011, *Sentiment analysis of consumer opinions written in Polish*, Economics and Management, tom 16, s. 1286–1291.
- Małyszko, J., 2013, *Aspect-Aware Identification of Opinion Phrases Polarity Based on Summaries of Consumer Opinions about Products and Services*, w: *Business Information Systems Workshops*, Springer, s. 278–289.
- Małyszko, J. i Filipowska, A., 2013, *Development of resources for the needs of sentiment analysis*, w: *Human Language Technologies as a Challenge for Computer Science and Linguistics. Proceedings of 6th Language and Technology Conference*, Fundacja Uniwersytetu im. A. Mickiewicza w Poznaniu (UAM) w Poznaniu, ISBN 98-83-932640-3-2, s. 381–385.
- Manning, C.D., Raghavan, P. i Schütze, H., 2008, *Introduction to information retrieval*, tom 1, Cambridge university press Cambridge.
- Maynard, D., Bontcheva, K. i Rout, D., 2012, *Challenges in developing opinion mining tools for social media*, Proceedings of@ NLP can u tag# usergeneratedcontent.

- Mayzlin, D., 2006, *Promotional chat on the Internet*, Marketing Science, s. 155–163.
- McAfee, A., 2006, *Enterprise 2.0: The dawn of emergent collaboration*, MIT Sloan Management Review, tom 47, nr 3, s. 21–28.
- Mihalcea, R., Banea, C. i Wiebe, J., 2007, *Learning multilingual subjective language via cross-lingual projections*, w: *Annual Meeting – Association for Computational Linguistics*, tom 45, s. 976.
- Miller, G.A., 1995, *WordNet: a lexical database for English*, Commun. ACM, tom 38, nr 11, s. 39–41, ISSN 0001-0782.
- Mitchell, T. i in., 1997, *Machine learning*.
- Moghaddam, S. i Ester, M., 2012, *Aspect-based Opinion Mining from Online Reviews*, http://www.sfu.ca/~sam39/papers/SIGIR_2012_Tutorial_Slides.pdf.
- Mudambi, S. i Schuff, D., 2010, *What makes a helpful online review? A study of customer reviews on Amazon. com*, MIS quarterly, tom 34, nr 1, s. 185–200.
- Mullen, T. i Malouf, R., 2006, *A preliminary investigation into sentiment analysis of informal political discourse*, w: *AAAI symposium on computational approaches to analysing weblogs (AAAI-CAAW)*, s. 159–162.
- Nasukawa, T. i Yi, J., 2003, *Sentiment analysis: capturing favorability using natural language processing*, w: *Proceedings of the 2nd international conference on Knowledge capture, K-CAP '03*, ACM, New York, NY, USA, ISBN 1-58113-583-1, s. 70–77.
- Ng, H.T. i Lee, H.B., 1996, *Integrating multiple knowledge sources to disambiguate word sense: An exemplar-based approach*, w: *Proceedings of the 34th annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics, s. 40–47.
- Niezurawski, L., Pawłowska, B. i Witkowska, J., 2010, *Satysfakcja klienta. Strategia – Pomiar – Zarządzenie. Koncepcja wewnętrznego urynkowania współczesnej organizacji*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń.
- Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A. i Sinz, E.J., 2011, *Memorandum on design-oriented information systems research*, European Journal of Information Systems, tom 20, nr 1, s. 7–10.

- Ott, M., Cardie, C. i Hancock, J., 2012, *Estimating the prevalence of deception in online review communities*, w: *Proceedings of the 21st international conference on World Wide Web*, ACM, s. 201–210.
- O'Reilly, T., 2004, *The Architecture of Participation*, http://www.oreillyn.com/pub/a/oreilly/tim/articles/architecture_of_participation.html, Dostęp 4 września 2012.
- O'Reilly, T., 2005, *What is Web 2.0*, <http://oreilly.com/web2/archive/what-is-web-20.html>, Dostęp 4 września 2012.
- O'Reilly, T., 2006, *Web 2.0 Compact Definition: Trying Again*, <http://radar.oreilly.com/2006/12/web-20-compact-definition-tryi.html>, Dostęp 4 września 2012.
- Panek, E., 2000, *Ekonomia matematyczna*, Wydawnictwo Akademii Ekonomicznej.
- Pang, B. i Lee, L., 2008, *Opinion mining and sentiment analysis*, Foundations and Trends in Information Retrieval, tom 2, nr 1-2, s. 1–135.
- Pang, B., Lee, L. i Vaithyanathan, S., 2002, *Thumbs up?: sentiment classification using machine learning techniques*, w: *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, Association for Computational Linguistics, s. 79–86.
- Piasecki, M., 2007, *Polish Tagger TaKIPI: Rule Based Construction and Optimisation*, Task Quarterly, tom 11, nr 1–2, s. 151–167.
- Pigla, W., 2012, *Webplemię – próba eksplikacji pojęcia*, Kultura i Historia, tom 21, ISSN 1642-9826.
- Popescu, A.M. i Etzioni, O., 2005, *Extracting product features and opinions from reviews*, w: *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, Association for Computational Linguistics, Stroudsburg, PA, USA, s. 339–346.
- Przepiórkowski, A., 2007, *Slavonic information extraction and partial parsing*, w: *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies*, Association for Computational Linguistics, s. 1–10.

- Qiu, G., Liu, B., Bu, J. i Chen, C., 2009, *Expanding Domain Sentiment Lexicon Through Double Propagation*, w: *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, IJCAI'09, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, s. 1199–1204, URL <http://dl.acm.org/citation.cfm?id=1661445.1661637>.
- Rekowski, M., 2008, *Mikroekonomia*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań.
- Rill, S., Scheidt, J., Drescher, J., Schütz, O., Reinel, D. i Wogenstein, F., 2012, *A generic approach to generate opinion lists of phrases for opinion mining applications*, w: *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, WISDOM '12, ACM, New York, NY, USA, ISBN 978-1-4503-1543-2, s. 7:1–7:8, URL <http://doi.acm.org/10.1145/2346676.2346683>.
- Rudnicki, L., 2004, *Zachowania rynkowe nabywców: mechanizmy i uwarunkowania*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- Sangiorgi, P., Augello, A. i Pilato, G., 2014, *An approach to detect polarity variation rules for sentiment analysis*, w: *Proc of WEBIST 2014–10th international conference on web information systems and technologies, 3–5 April*.
- Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H. i Jin, C., 2007, *Red Opal: product-feature scoring from reviews*, w: *Proceedings of the 8th ACM conference on Electronic commerce*, EC '07, ACM, New York, NY, USA, ISBN 978-1-59593-653-0, s. 182–191.
- Sebastiani, F., 2002, *Machine learning in automated text categorization*, ACM Comput. Surv., tom 34, nr 1, s. 1–47, ISSN 0360-0300, URL <http://doi.acm.org/10.1145/505282.505283>.
- Shapiro, C., 1982, *Consumer information, product quality, and seller reputation*, The Bell Journal of Economics, s. 20–35.
- Sharma, A. i Dey, S., 2012, *A Comparative Study of Feature Selection and Machine Learning Techniques for Sentiment Analysis*, w: *Proceedings of the 2012 ACM Research in Applied Computation Symposium*, RACS '12, ACM, New York, NY, USA, ISBN 978-1-4503-1492-3, s. 1–7, URL <http://doi.acm.org/10.1145/2401603.2401605>.
- Sharma, A. i Dey, S., 2013, *A Boosted SVM Based Sentiment Analysis Approach for Online Opinionated Text*, w: *Proceedings of the 2013 Research in Adaptive and Convergent Systems*,

- RACS '13, ACM, New York, NY, USA, ISBN 978-1-4503-2348-2, s. 28–34, URL <http://doi.acm.org/10.1145/2513228.2513311>.
- Silva, M., Carvalho, P. i Sarmiento, L., 2012, *Building a Sentiment Lexicon for Social Judgement Mining*, w: H. Caseli, A. Villavicencio, A. Teixeira i F. Perdigão (red.), *Computational Processing of the Portuguese Language*, tom 7243 z serii *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, ISBN 978-3-642-28884-5, s. 218–228.
- Smallwood, D.E. i Conlisk, J., 1979, *Product quality in markets where consumers are imperfectly informed*, *The Quarterly Journal of Economics*, s. 1–23.
- Smeureanu, I. i Bucur, C., 2012, *Applying supervised opinion mining techniques on online user reviews*, *Informatica Economica*, tom 16, nr 2, s. 81–91.
- Spence, M., 1973, *Job market signaling*, *The quarterly journal of Economics*, tom 87, nr 3, s. 355–374.
- Stone, B. i Richtel, M., 2007, *The Hand That Controls the Sock Puppet Could Get Slapped*, *The New York Times*, <http://www.nytimes.com/2007/07/16/technology/16blog.html?pagewanted=all>, dostep 10.09.2012.
- Su, X. i Khoshgoftaar, T.M., 2009, *A survey of collaborative filtering techniques*, *Adv. in Artif. Intell.*, tom 2009, s. 4:2–4:2, ISSN 1687-7470, URL <http://dx.doi.org/10.1155/2009/421425>.
- Subasic, P. i Huettner, A., 2001, *Affect analysis of text using fuzzy semantic typing*, *Fuzzy Systems, IEEE Transactions on*, tom 9, nr 4, s. 483–496.
- Sun, J., Long, C., Zhu, X. i Huang, M., 2009, *Mining reviews for product comparison and recommendation*, *Polibits*, tom 39, s. 33–40.
- Suresh, V., Roohi, S. i Eirinaki, M., 2014, *Aspect-based Opinion Mining and Recommendation-system for Restaurant Reviews*, w: *Proceedings of the 8th ACM Conference on Recommender Systems*, RecSys '14, ACM, New York, NY, USA, ISBN 978-1-4503-2668-1, s. 361–362, URL <http://doi.acm.org/10.1145/2645710.2645716>.
- Szwed, R., 2009, *Metody statystyczne w naukach społecznych. Elementy teorii i zadania*, Katolicki Uniwersytet Lubelski, Lublin.

- Taboada, M., Brooke, J., Tofiloski, M., Voll, K. i Stede, M., 2011, *Lexicon-based methods for sentiment analysis*, Computational linguistics, tom 37, nr 2, s. 267–307.
- Takamura, H., Inui, T. i Okumura, M., 2006, *Latent variable models for semantic orientations of phrases*, w: *Proceedings of the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL 2006)*.
- The Nielsen Company, 2009, *Personal Recommendations And Consumer Opinions Posted Online Are The Most Trusted Forms Of Advertising Globally*, http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/07/pr_global-study_07709.pdf.
- Thet, T., Na, J. i Khoo, C., 2010, *Aspect-based sentiment analysis of movie reviews on discussion boards*, Journal of Information Science, tom 36, nr 6, s. 823–848.
- Turney, P.D., 2002, *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews*, w: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, Association for Computational Linguistics, Stroudsburg, PA, USA, s. 417–424.
- Turney, P.D. i Littman, M.L., 2003, *Measuring praise and criticism: Inference of semantic orientation from association*, ACM Trans. Inf. Syst., tom 21, nr 4, s. 315–346, ISSN 1046-8188.
- Wan, X., 2009, *Co-training for cross-lingual sentiment classification*, w: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, Association for Computational Linguistics, s. 235–243.
- Wang, H., Lu, Y. i Zhai, C., 2010, *Latent aspect rating analysis on review text data: a rating regression approach*, w: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, s. 783–792.
- Wawer, A., 2011, *Mining opinion attributes from texts using multiple kernel learning*, w: *IEEE 11th International Conference on Data Mining Workshops*.
- Wawer, A., 2012, *Mining Co-Occurrence Matrices for SO-PMI Paradigm Word Candidates*, w: *Proceedings of 13th Conference of the European Chapter of the Association for computational Linguistics (EACL 2012): Student Research Workshop*, forthcoming.

- Wieloch, K., 2011, *Budowanie profili przedsiębiorstw z wykorzystaniem utożsamiania odwołań sematycznych do przedsiębiorstw polskich w polskich tekstach ekonomicznych*, Rozprawa doktorska, "Uniwersytet Ekonomiczny w Poznaniu".
- Wilson, T., Wiebe, J. i Hoffmann, P., 2005, *Recognizing contextual polarity in phrase-level sentiment analysis*, w: *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, Association for Computational Linguistics, Stroudsburg, PA, USA, s. 347–354, URL <http://dx.doi.org/10.3115/1220575.1220619>.
- Yu, J., Zha, Z., Wang, M. i Chua, T., 2011, *Aspect ranking: identifying important product aspects from online consumer reviews*, *Computational Linguistics*, s. 1496–1505.
- Zhang, P.F., Su, Y.J. i Wang, C., 2007, *Statistical Machine Learning Used in Integrated Anti-Spam System*, w: *Machine Learning and Cybernetics, 2007 International Conference on*, tom 7, s. 4055–4058.
- Zhu, F. i Zhang, X., 2010, *Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics*, *Journal of Marketing*, tom 74, nr 2, s. 133–148.

Załączniki

A Adnotacja opinii wyrażanych przez recenzentów w dokumentach tekstowych – wskazówki adnotacji

A.1 Cel procesu adnotacji

Celem adnotacji jest przygotowanie przetwarzalnych reprezentacji opinii wyrażonych w wypowiedziach tekstowych przez recenzentów. W takich wypowiedziach recenzenci opisują swoje opinie o recenzowanych produktach na różne sposoby. Nadawane adnotacje mają na celu dostarczenie informacji o takich opiniach wyrażonych w tekście.

Zgromadzone adnotacje pozwolą na przeprowadzenie analizy spójności pomiędzy opiniami wyrażonymi przez recenzentów w wypowiedziach tekstowych oraz oznaczonych przez nich w listach zalet i wad przypisanych do danej recenzji.

A.2 Typy adnotacji

Adnotatorzy mieli za zadanie oznaczanie opinii recenzentów o aspektach omówionych w tabeli 1 na stronie 181. Wszelkie opinie dotyczące innych aspektów powinny być przez adnotatorów ignorowane. Każdy z aspektów recenzenci mogli ocenić na jeden z dwóch sposobów: pozytywnie lub negatywnie. W związku z tym, do każdego aspektu istnieją dwa typy adnotacji: odpowiadający uznaniu danego aspektu za zaletę lub za wadę recenzowanego produktu. Wszystkie typy adnotacji zostały zaprezentowane w tabeli 2 na stronie 186.

A.3 Wskazówki adnotacji

Adnotacje przeprowadzane będą z wykorzystaniem środowiska GATE¹. W trakcie adnotacji zadanie adnotatora polega na czytaniu zadanych recenzji tekstowych i identyfikacji aspektów, o

¹<http://gate.ac.uk/>

których recenzent wyraził swoją opinię i jaka to była opinia (tj. pozytywna czy negatywna). Lista zalet i wad nadana przez recenzenta jest przed adnotatorem zakryta.

Przykładowo, widząc zdanie „Aparat foto 5mpx z lampą błyskowa robi fenomenalne zdjęcia, ale wbudowany wyświetlacz ma bardzo kiepskie odwzorowanie kolorów.”, adnotator powinien przypisać adnotację: *dobry aparat fotograficzny*. W zdaniu pojawia się dodatkowo ocena odwzorowania kolorów przez wyświetlacz, jednak nie zdefiniowano odpowiedniego aspektu w tabeli 1, w związku z tym adnotator powinien zignorować tę opinię.

W trakcie adnotacji należy kierować się następującymi wskazówkami:

- za każdym razem, gdy adnotator uważa, że z tekstu jasno wynika opinia recenzenta o danym aspekcie, należy nadać adnotację, niezależnie od tego, w jaki sposób została ona wyrażona,
- należy nadawać adnotację tylko wówczas, gdy oceniany jest aspekt recenzowanego produktu; w recenzjach pojawiają się również opinie o innych produktach, które powinny być ignorowane,
- w recenzjach adnotatorzy mogą napotkać opinie porównawcze, tj. takie, gdzie dany produkt jest oceniany w porównaniu do produktu innego. W takich sytuacjach często nie ma wyrażonej opinii bezwzględnej (np. że dany aspekt podoba się recenzentowi), a jedynie względna, tj. że pewien aspekt w recenzowanym produkcie jest lepszy niż w produkcie innym. W takich sytuacjach może nie być jasne, czy dany aspekt jest faktycznie zaletą produktu (być może to, że dany aspekt jest lepszy w recenzowanym produkcie niż w produkcie innym nie oznacza, że jest faktycznie zaletą produktu). W takim przypadku adnotatorzy na podstawie analizy szerszego kontekstu wypowiedzi (np. wcześniejszych i kolejnych zdań) powinni ustalić, czy należy nadać adnotację,
- w trakcie adnotacji nie jest istotny zakres adnotacji, tj. jaki dokładnie fragment tekstu został zaadnotowany, bowiem w dalszym przetwarzaniu ta informacja nie będzie wykorzystywana. Adnotatorzy powinni orientacyjnie oznaczyć, w którym fragmencie tekstu wyrażona została opinia, ale nie jest to sprawa kluczowa dla wykorzystania efektów adnotacji

B Budowanie leksykonu nazw aspektów na podstawie ręcznej analizy korpusu – wskazówki adnotacji

B.1 Cel procesu adnotacji

Celem procesu adnotacji jest ręczne skonstruowanie leksykonu sformułowań będących nazwami aspektów produktów ocenianych w analizowanym korpusie, tj. telefonów komórkowych. Adnotatorzy mają za zadanie przeanalizować korpus dokumentów tekstowych, będących recenzjami konsumentekimi.

Recenzje wylosowano z trzech wybranych portali (`cokupic.pl`, `opineo.pl` i `bozo.pl`). W sumie liczba recenzji wynosi 450 (po 150 recenzji dla każdego z wymienionych portali, przy czym z portalu `bozo.pl` 50 dokumentów to recenzje tekstowe, 50 to tekst wprowadzony przez użytkownika do listy zalet i 50 to tekst wprowadzony do listy wad). W każdej recenzji może pojawić się wiele odniesień do wielu aspektów. Zadaniem adnotatorów jest przeczytanie recenzji i w każdej z nich zaznaczenie wyrazów (lub dłuższych sformułowań), które ich zdaniem są nazwami aspektów produktów pochodzących z podanej listy. Przez **aspekt produktu**, na obecnym etapie, za pracami [Liu, Hu i Cheng 2005; Sun i in. 2009] rozumiemy jego cechę (ang. *feature*), która rozumiana jest jako atrybut bądź komponent produktu, do którego recenzenci odnoszą się w swoich opiniach. Natomiast **nazwa aspektu produktu** jest to rzeczownik lub fraza go zawierająca używana przez recenzentów w celu odniesienia się w tekście do danego aspektu.

Każdy aspekt posiada jedną domyślną, ustaloną z góry nazwę. W recenzjach mogą pojawiać się inne wyrazy lub frazy, za pomocą których recenzenci odnoszą się do danego aspektu. Znajdując takie wyrazy, adnotatorzy mają je oznaczyć za pomocą domyślnej nazwy danego aspektu. Innymi słowy, w trakcie procesu adnotacji mamy do czynienia z różnymi **typami adnotacji**. Przykładowo, jeśli domyślną nazwą jednego ze zdefiniowanych aspektów produktu jest *waga*, inną nazwą aspektu może być np. *ciężar*. W związku z tym adnotujemy ten wyraz za pomocą schematu adnotacji *waga*.

Wyrazy oznaczone przez adnotatorów zostaną następnie automatycznie pobrane z korpusu i pogrupowane według aspektu, którego dotyczą. Aby zapewnić, że wygenerowany w ten sposób leksykon będzie jak najwyższej jakości, adnotacja przeprowadzona zostanie przez dwie osoby na podstawie ustalonych wskazówek adnotacji. Ustalona zostanie zgodność pomiędzy adnotacjami nadanymi przez obu adnotatorów.

B.2 Aspekty i typy adnotacji

W tabeli 1 zaprezentowano listę aspektów, których nazwy odnalezione w tekście mają być oznaczane przez adnotatorów. Dla każdego aspektu istnieje jeden typ adnotacji posiadający analogiczną nazwę, gdzie w celu uniknięcia błędów w trakcie późniejszego automatycznego przetwarzania w nazwach typów adnotacji polskie znaki zamieniono na ich odpowiedniki z alfabetu łacińskiego, zaś spacje zamieniono na podkreślniki.

aspekt	wyjaśnienie
aparat fotograficzny	aparat fotograficzny wbudowany w telefon
awaryjność	jak często telefon ulega awariom wymuszającym na użytkowniku zanieśenie urządzenia do serwisu
cena	cena, w jakiej produkt jest dostępny na rynku i jej stosunek do jakości produktu lub usługi
design	na ile wygląd telefonu jest atrakcyjny dla użytkowników (czy telefon się użytkownikowi podoba wizualnie); chodzi tu o ogólne subiektywne postrzeganie wyglądu całego urządzenia przez recenzenta
gps	jakość wbudowanego modułu nawigacji satelitarnej (czy spełnia swoją rolę)
liczba funkcji	jak wiele telefon ma funkcji; nazwami aspektu powinny być wyłącznie określenia opisujące ogólną liczbę funkcji udostępnianych przez telefon, a nie wyliczenia poszczególnych funkcji dostępnych w telefonie (przykładowo, fraza „moduł wifi” nie powinna być zaadnotowana, ponieważ pomimo tego, że jeśli jest dostępny to zwiększa liczbę funkcji, to nie opisuje bezpośrednio liczby dostępnych funkcji i być może oprócz tego modułu telefon posiada bardzo niewiele innych funkcji)
jakość dźwięku	jakość dźwięku podczas rozmowy, brzmienie odtwarzanej muzyki (np. przez radio i odtwarzacz mp3), również w trybie głośnomówiącym oraz na słuchawkach
kompatybilność	kompatybilność telefonu z innymi urządzeniami elektronicznymi i oprogramowaniem

aspekt	wyjaśnienie
obsługa	prostota obsługi telefonu zarówno od strony sprzętowej (np. rozplanowanie przycisków), jak i programowej (przejrzystość menu itp)
odtwarzacz mp3	funkcja telefonu odtwarzająca nagrania dźwiękowe zapisane w plikach mp3 i podobnych (np. ogg)
parametry	parametry techniczne, którymi cechuje się telefon; aspekt ten odnosi się do podzespołów telefonu wpływających na jego pracę, które nie mieszczą się w pozostałych aspektach wymienionych w tabeli
pojemność	aspekt określa, jak wiele danych (np. wiadomości, filmów, aplikacji) jest w stanie pomieścić pamięć wbudowana w urządzenie
rozdzielczość wyświetlacza	liczba pikseli obrazu wyświetlanych na wyświetlaczu urządzenia; powiązana z takimi odczuciami recenzentów, jak czytelność obrazu (na ile obrazy, filmy czy litery wyświetlane na ekranie są wyraźne)
rozmiar wyświetlacza	na ile rozmiar wyświetlacza, w który urządzenie jest wyposażone, pozwala na komfortowe korzystanie z urządzenia
serwis	jakość serwisu zapewnianego przez producenta urządzenia
software	liczba i jakość aplikacji zainstalowanych na telefonie bądź możliwych do pobrania i uruchomienia na nim
sygnał	na ile dobrze telefon utrzymuje połączenie z siecią operatora
szybkość	odnosi się do czasu reakcji urządzenia na polecenia wydawane przez użytkownika (np. średnia długość okresów wczytywania kolejnych list menu itp.)
waga	czy ciężar telefonu pozwala na swobodne korzystanie z urządzenia
wymiary	czy wymiary telefonu pozwalają na swobodne korzystanie z urządzenia
wytrzymałość	w jakim stopniu telefon jest odporny na uszkodzenia mechaniczne wynikające z kontaktu z otoczeniem (np. pęknięcia obudowy, powstawanie rys na wyświetlaczu itp.)
żywość baterii	czy okres pomiędzy ładowaniami baterii umożliwia bezproblemowe korzystanie z urządzenia

Tabela 1: Lista aspektów produktów (telefonów komórkowych), dla których w wyniku procesu adnotacji wygenerowane mają zostać słowniki nazw. Źródło: opracowanie własne na podstawie strony cokupic.pl

B.3 Wskazówki adnotacji

Odnoszenie się do aspektu wprost i nie wprost

Aspekt może być wspomniany w teście wprost i nie wprost. Mówiąc, że aspekt jest wspomniany w tekście wprost mamy na myśli, że pojawia się tam jego nazwa. Natomiast aspekt wspomniany jest nie wprost wówczas, gdy w tekście pojawia się sformułowanie, które bezpośrednio dotyczy tego aspektu, jednak nigdzie nie pojawia się jego nazwa. Przykładowo, w zdaniu „Telefon jest tani” występuje odniesienie do aspektu *cena*. Odniesienie to odbywa się tu poprzez słowo „tani”, które jednak nie jest rzeczownikiem ani frazą zawierającą rzeczownik. W związku z tym nie występuje tutaj nazwa aspektu zgodna z definicją podaną w sekcji C.1 i adnotacja nie powinna zostać nadana. Podsumowując, jeśli w tekście pojawi się odniesienie do aspektu nie wprost, należy nie nadawać żadnej adnotacji.

Ustalanie typu adnotacji

Istotną rolę adnotatora jest określenie, czy dane sformułowanie jest nazwą któregoś z aspektów wymienionych w tabeli 1 i jeśli tak, to którego. Istotne jest to, aby nazwa była nadawana tylko wtedy, gdy analizowane zdanie jest zgodne z wyjaśnieniem danego aspektu zawartym w drugiej kolumnie tabeli. Może się okazać, że pewne sformułowanie jest powiązane z danym aspektem, ale w przekonaniu adnotatora nie można jednak go utożsamić znaczeniowo z danym aspektem. Adnotatorzy powinni kierować się tutaj dwiema podstawowymi regułami:

- jeśli ocena aspektu opisywanego przez typ adnotacji oraz przez analizowane sformułowanie mogą być rozbieżne, to dane sformułowanie nie jest nazwą aspektu. Przykładowo, słowo „głośność” w przypadku telefonu komórkowego jest powiązane z aspektem *jakość dźwięku*. Jednak możliwe jest, że recenzent pozytywnie oceni głośność, jednak pomimo tego jakość dźwięku może być oceniona słabo (np. dźwięki wydawane przez głośniki telefonu są donośne, ale brzmią sztucznie). W takim przypadku, „głośność” nie powinna być zaadnotowana jako nazwa aspektu *jakość dźwięku*,
- potencjalna nazwa aspektu, po połączeniu z przymiotnikami „dobry” lub „zły” powinna tworzyć frazę, która wyraźnie wskazuje na to, jakiego aspektu dotyczy nazwa. Przykładowo, w sformułowaniu „ekran szybko się rysuje” adnotator zastanawia się, czy „ekran ” można zaadnotować jako nazwę aspektu *wytrzymałość* (gdyż cała fraza dotyczy tego aspektu, a „ekran” jest w niej jedynym rzeczownikiem). Jednakże, jeśli skonstruujemy frazę „dobry

ekran”, to otrzymana fraza jest odległa znaczeniowo od aspektu *wytrzymałość* i prawdopodobnie częściej kojarzy się z tym, ekran ma np. duży rozmiar, wysoką rozdzielczość i dobre odwzorowanie kolorów. W takim przypadku adnotator nie powinien nadawać adnotacji.

Możliwe jest zaistnienie sytuacji, w których problematyczne będzie ustalenie typu adnotacji, który powinien zostać nadany (to jest ustalenie aspektu, do którego występuje odniesienie w analizowanym zdaniu). Przykładowo, słowo „obudowa” w sformułowaniu „piękna obudowa” odnosi się do aspektu *design*, natomiast w „niezniszczalna obudowa” do aspektu *wytrzymałość*. W takich przypadkach należy nadawać adnotację wynikającą z kontekstu. Innymi słowy, wyraz „obudowa” może mieć nadane adnotacje różnych typów w obu wspomnianych przypadkach.

Możliwe jest również wystąpienie słowa, które w danym kontekście może się odnosić do więcej niż jednego aspektu. W takim przypadku, adnotatorzy powinni nadać kilka adnotacji.

Zakres adnotacji

Problematycznym zadaniem w niektórych przypadkach może być ustalenie zakresu adnotacji, tj. określenie jak wiele słów dana adnotacja powinna obejmować - czy powinien być to tylko jeden wyraz, czy też może dłuższa fraza – a jeśli fraza, to jak ma być długa. Przy ustalaniu długości frazy należy wziąć pod uwagę następujące reguły:

- adnotacja powinna obejmować wszystkie słowa, które wymagane są do tego, aby poprawnie zrozumieć sens zdania, w którym wystąpiło w tekście. Przykładowo, w zdaniu „Czas czuwania jest bardzo krótki” należy zaadnotować frazę „czas czuwania” jako nazwę aspektu *żywołność baterii*, gdyż samo słowo „czas” nie zachowuje tutaj właściwego znaczenia i po usunięciu słowa „czuwania” nie bylibyśmy w stanie odgadnąć, jakiego aspektu dane słowo się dotyczy. Przykładowo, samo słowo „czas” może odnosić się w kontekście telefonu komórkowego do czasu włączania się urządzenia lub czasu ładowania się baterii. W związku z tym, należy zaadnotować w tym przypadku całą frazę „czas czuwania”. Podobnie, w zdaniu „Wadą jest błyskawiczne rozładowywanie się baterii”, nazwą aspektu powinno być całe sformułowanie „rozładowywanie się baterii”, gdyż usunięcie słowa „rozładowywanie się” nie pozwala na zrozumienie sensu zdania.
- fraza będąca nazwą aspektu nie może zawierać żadnych słów, które opisują dany aspekt w kontekście konkretnego, opisywanego produktu. Przykładowo, w zdaniu „Aluminiowa obudowa jest bardzo wytrzymała” zaadnotowanie frazy „Aluminiowa obudowa” byłoby błędne,

gdyż to ten konkretny model telefonu ma aluminiową obudowę; należy w tym przypadku jako nazwę aspektu oznaczyć wyłącznie słowo „obudowa”.

- nazwy aspektu nie mogą wskazywać w żaden sposób na opinię autora tekstu o danym aspekcie. Przykładowo, fraza *piękna obudowa* byłaby błędną adnotacją.

C Budowanie bazy przykładowych sformułowań używanych do oceny aspektów produktów – wskazówki adnotacji

C.1 Cel procesu adnotacji

Recenzenci wyrażają swoje opinie o produktach za pomocą różnego rodzaju sformułowań. Celem procesu adnotacji jest zebranie bazy przykładowych sformułowań wykorzystywanych przez recenzentów dla określania opinii. Adnotatorzy mają za zadanie oznaczyć frazy, w których oceniany jest któryś z aspektów zdefiniowanych w tabeli 2, przy czym istotne są tu tylko te sformułowania, w których odnoszono się do aspektu wprost (czyli w których pojawiła się nazwa aspektu).

Zgromadzona baza sformułowań, za pomocą których recenzenci oceniali aspekty produktów, zostanie następnie wykorzystana do przygotowania reguł ekstrakcji leksykalnej, dzięki którym możliwe będzie automatyczne identyfikowanie fragmentów tekstu, w których recenzenci oceniają aspekty produktów.

C.2 Typy adnotacji i powiązania z aspektami produktów

W tabeli 2 zaprezentowano typy adnotacji, za pomocą których adnotatorzy powinni oznaczać fragmenty tekstów, w których oceniane są poszczególne aspekty. Kolumny w tej tabeli mają następującą interpretację:

- wypisane w pierwszej kolumnie aspekty mają interpretację identyczną, jak w tabeli 1 na stronie 181,
- w drugiej kolumnie znajdują się typy adnotacji, za pomocą których adnotatorzy mają oznaczać sformułowania, w których recenzenci pozytywnie ocenili dany aspekt; w programie GATE każdy z tych typów adnotacji ma dodatkowo nadany przedrostek *zaleta*,
- w trzeciej kolumnie wypisano typy adnotacji oznaczające, że zaadnotowane za ich pomocą fragmenty tekstu zawierają negatywną ocenę danego aspektu; w GATE dodano do nich przedrostek *wada*.

Oprócz wspomnianych przedrostków do typów adnotacji, w programie GATE poszczególne typy adnotacji dodatkowo mają zamienione polskie litery na ich łańciskie odpowiedniki (np. litera *ś* została wszędzie zamieniona na *s* itp.) oraz spacje zamieniono na znaki podkreślnika. W rezultacie, przykładowo typ adnotacji oznaczający pozytywną ocenę aspektu *aparat fotograficzny* w programie GATE jest dostępny jako *zaleta_dobry_aparat_fotograficzny*.

aspekt	zaleta	wada
aparat fotograficzny	dobry aparat fotograficzny	kiepski aparat fotograficzny
awaryjność	niezawodność	awaryjność
cena	cena	cena
design	atrakcyjny design	nieatrakcyjny design
gps	funkcjonalny gps	słaby/brak gps
liczba funkcji	dużo funkcji	brak dodatkowych funkcji
jakość dźwięku	optymalna wyrazistość dźwięku	słaba wyrazistość dźwięków
obsługa	łatwość obsługi	skomplikowana obsługa
odtwarzacz mp3	świetny odtwarzacz mp3	brak/kiepski odtwarzacz mp3
parametry	świetne parametry	brak dostatecznych parametrów
pojemność	pojemność	pojemność
rozdzielczość wyświetlacza	dobra rozdzielczość	niska rozdzielczość wyświetlacza
rozmiar wyświetlacza	duży wyświetlacz	mały wyświetlacz
software	software	słaby software
sygnał	silny sygnał	słaby sygnał
szybkość	szybkość	wolny
waga	lekkość	ciężki
wymiary	odpowiednie wymiary	kiepskie wymiary
wytrzymałość	wytrzymałość	delikatna konstrukcja
żywność baterii	długa żywotność baterii	słaba bateria

Tabela 2: Lista aspektów produktów (telefonów komórkowych) wraz z odpowiadającymi im typami adnotacji oznaczającymi pozytywne lub negatywne postrzeganie danego aspektu przez recenzenta. Źródło: opracowanie własne na podstawie strony cokupic.pl.

C.3 Wskazówki adnotacji

Każda adnotacja powinna zawierać dwa główne elementy: nazwę ocenianego aspektu oraz słowo lub dłuższe sformułowanie wyrażające opinię o danym aspekcie. Poniżej omówiono podstawowe reguły, którymi powinni się kierować adnotatorzy.

- Adnotacja powinna zawierać wystarczający zakres tekstu, aby po przeczytaniu jej samej (bez czytania kontekstu, w którym się znajduje) było jednoznaczne, jaka jest opinia recenzenta o danym aspekcie. Przykładowo, w zdaniu „Telefon charakteryzuje się długo trzymającą baterią”, w którym nazwą aspektu jest słowo „baterią”, należy zaadnotować frazę „długo trzymającą baterią”. Te trzy wyrazy wystarczą, aby poznać opinię recenzenta.
- Adnotatorzy powinni oznaczać wyłącznie frazy, w których recenzenci wyrażają swoją opinię o aspektach wprost, tj. używając ich nazwy. Przykładem wyrażenia opinii wprost jest fraza „niska waga”, natomiast nie wprost jest słowo „lekki”; oba te sformułowania dotyczą aspektu *waga*, jednak nazwa aspektu pojawia się tylko w pierwszym z nich i tylko takie frazy mają zaznaczać adnotatorzy. Adnotacja przeprowadzana jest na dokumentach z wcześniej ręcznie zaadnotowanymi nazwami aspektów. W związku z tym, każda adnotacja nadana w tej fazie powinna zawierać w sobie adnotację nazwy aspektu z poprzedniej fazy adnotacji (tzn. zakres tekstu, który oznaczają adnotatorzy, powinien obejmować sformułowanie oznaczone jako nazwa aspektu).
- Adnotatorzy mogą natknąć się na zdania, w których znajduje się przeczenie dotyczące frazy wyrażającej opinie. Przykładowo, w zdaniu „Bateria nie jest wystarczająca” lub „Nie zgadzam się ze stwierdzeniem że cena jest wysoka” takim przeczącym słowem jest „nie”. Podobną funkcję w zdaniu mogą pełnić przykładowo takie słowa jak „nigdy”, „żaden” itp. Reguła postępowania w takim przypadku jest taka, że powinna być nadawana adnotacja taka, jaka byłaby nadana gdyby fraza nie była zanegowana, jednak dodatkowo oznaczamy słowo negujące za pomocą specjalnego typu adnotacji *negacja*. Możemy mieć do czynienia z dwoma przypadkami:
 - słowo powodujące negację znajduje się pomiędzy nazwą aspektu a słowem wyrażającym opinię; w takim przypadku adnotacja frazy wyrażającej opinie powinna zawierać w sobie to słowo, które dodatkowo należy oznaczyć za pomocą adnotacji typu *negacja* (nadajemy dwie adnotacje, gdzie jedna znajduje się wewnątrz drugiej),

- negacja znajduje przed lub po fragmencie zdania, w którym znajdują się nazwa aspektu i słowo wyrażające opinię. W takim przypadku, należy nadać dwie rozłączne adnotacje; jedną oznaczającą słowo będące negacją, a drugą oznaczającą frazę wyrażającą zanegowaną opinię.
- oprócz negacji, problematycznymi sytuacjami mogą być zdania warunkowe. Przykładowo, w zdaniu „Wyświetlacz mógłby być większy” czy też „Wyświetlacz powinien być większy”, aspekt o nazwie „rozmiar wyświetlacza” opisywany jest za pomocą słowa „większy”, jednak dodatkowo mamy tu do czynienia ze słowami „mógłby” i „powinien”, zmieniającymi sens wypowiedzi. Ponownie, adnotujemy takie zdanie tak jakby nie było słowa zmieniającego sens, ale nadajemy dodatkową adnotację o specjalnym typie *zmiana* danemu słowu zmieniającemu sens wypowiedzi.
 - Słowa użyte przez adnotatorów dzielą się na dwa rodzaje: uniwersalne i specyficzne dla aspektów. Uniwersalne słowa to takie, które niezależnie od tego, jakiego aspektu dotyczą, oznaczają taką samą opinię recenzenta o aspekcie (pozytywną lub negatywną). Najprostszymi przykładami takich słów są „dobry” i „zły”. Słowa specyficzne dla aspektów natomiast dla niektórych aspektów mogą oznaczać pozytywną opinię recenzenta (np. „niska cena”) a w innym przypadku negatywną (np. „niska jakość dźwięku”). Adnotatorzy mają adnotować frazy niezależnie od tego czy występują w nich słowa uniwersalne czy specyficzne dla aspektu.

D Oznaczanie poprawności wyekstrahowanych fraz używanych do oceny aspektów produktów – wskazówki adnotacji

D.1 Cel procesu adnotacji

Po przeprowadzeniu analizy bazy przykładowych sformułowań używanych do oceny aspektów produktów podjęto decyzję o podzieleniu procesu ekstrakcji fraz służących do wyrażania opinii na dwa etapy:

1. ekstrakcja fraz z wykorzystaniem niewielkiej liczby bardzo ogólnych reguł,
2. przefiltrowanie wyekstrahowanych fraz z wykorzystaniem statystycznej analizy kontekstu, w którym dana fraza się znajdowała.

Ekstrahowane frazy składały się z nazwy aspektu oraz sformułowania użytego do wyrażenia opinii o aspekcie. W niniejszym dokumencie wyekstrahowane pary będziemy zapisywać w znormalizowanej postaci za pomocą par (*sformułowanie, aspekt*), gdzie wszystkie słowa w sformułowaniu są sprowadzane do podstawowej formy gramatycznej, zaś aspekt jest reprezentowany przez swą podstawową nazwę, np. (*tatwy, obsługa*) czy (*wiele; ciekawy, liczba_funkcji*).

Pierwszy etap miał na celu ekstrakcję jak największej liczby fraz, za pomocą których recenzenci potencjalnie mogli oceniać aspekty produktów. Etap ten został zrealizowany z wykorzystaniem ręcznie przygotowanych reguł ekstrakcji, napisanych za pomocą formalizmu JAPE. Przygotowane reguły są wykorzystywane w potoku przetwarzania w pakiecie GATE, który za ich pomocą przetwarza dokumenty tekstowe z zadanego korpusu. Ekstrahowane są następujące dane o wyekstrahowanych fragmentach tekstu:

- jaki aspekt jest w nim oceniany oraz za pomocą jakiego sformułowania (para (*sformułowanie, aspekt*)),
- jak brzmi dokładna fraza, w której wykorzystano parę,
- jaki jest szerszy kontekst (ang. *snippet*) wyekstrahowanej frazy (okno składające się z wyrazów będących w odległości do 50 liter od frazy, maksymalnie do granicy zdania).

W drugiej fazie wyekstrahowane przykłady poddawane mają być analizie statystycznej w celu odrzucenia tych przykładów, w których prawdopodobnie nie poddawano ocenie aspektów produktu. W trakcie analizy wykorzystywane mają być takie statystyki, jak np. liczba przymiotników

w szerszym kontekście danego przykładu. Na podstawie takich statystyk, przygotowany zostanie klasyfikator, przypisujący każdą frazę do jednej z dwóch klas:

- zaakceptowane – frazy prawidłowo wyekstrahowane przez przygotowane reguły i które powinny zostać uwzględnione w dalszym przetwarzaniu,
- odrzucone – frazy, które nie będą dalej wykorzystywane w pracy.

Reguły, zgodnie z którymi adnotatorzy mają przypisywać przykłady do obu tych klas opisane są poniżej w podsekcji D.2. Wspomniany klasyfikator zostanie przygotowany z wykorzystaniem uczenia maszynowego, tj. na podstawie dostarczonych przykładów pochodzących z obu grup (zaakceptowane oraz odrzucone) w sposób automatyczny ² przygotowany zostanie model, który każdy nowy przykład będzie przypisywał do jednej ze wspomnianych klas.

D.2 Wskazówki adnotacji

Kwestie techniczne

Adnotacja przeprowadzana jest z wykorzystaniem arkusza kalkulacyjnego. Dane do adnotacji są udostępniane w pliku w formacie OASIS OpenDocument (ods), możliwym do edycji przykładowo za pomocą programu LibreOffice Calc w wersji 4.2.2.1. W pliku poszczególne wiersze odpowiadają przykładom wyekstrahowanym w fazie pierwszej. Kolumny mają następującą interpretację:

- (A) identyfikator przykładu,
- (B) aspekt oceniany w przykładzie,
- (C) znormalizowane sformułowanie, za pomocą którego oceniany jest aspekt (słowa sprowadzone są do podstawowej formy gramatycznej; jeśli jest ich więcej niż jedno, wypisywane są w nawiasie w kolejności, w której znalazły się w tekście),
- (D) fraza, w której pojawiła się para złożona z wartości w kolumnach B oraz C,
- (E) szerszy kontekst frazy; fraza została w nim wyszczególniona w następujący sposób: >>>fraza<<<<,
- (F) kolumna, w której mają znajdować się adnotacje.

W kolumnie F w każdej linii adnotator ma nadać jedną z dwóch wartości, odpowiadających klasom opisanym w punkcie D.1:

²Wykorzystane zostanie tu jedno z istniejących podejść, np. naiwny klasyfikator Bayesa.

1 – przykład powinien trafić do kategorii zaakceptowanych,

0 – przykład należy zaliczyć do kategorii odrzuconych.

Reguły przypisywania przykładów do klas

Każdy przykład powinien być przeanalizowany przez adnotatora i przypisany do odpowiedniej klasy zgodnie z poniższymi regułami:

- zaakceptowane – do tej klasy powinny zostać zaklasyfikowane przykłady spełniające następujące warunki:
 1. sformułowanie z wyekstrahowanej pary (*sformułowanie, aspekt*) może zostać użyte do wyrażenia opinii o konkretnym aspekcie (pozytywnej, negatywnej lub neutralnej) i nawet bez analizy kontekstu adnotator jest w stanie łatwo przewidzieć, jaką opinię dane sformułowanie może wyrażać (pozytywną, negatywną lub neutralną),
 2. z analizy szerszego kontekstu danego przykładu wynika, że recenzent w danym przykładzie wygłasza opinię o aspekcie z pary (*sformułowanie, aspekt*) zgodnie z wydźwiękiem, jaki zostałby przypisany do pary zgodnie z regułą 1,
- odrzucone – w tej klasie mają się znaleźć przykłady, które nie spełniają co najmniej jednej z reguł sformułowanych w poprzednim punkcie.

W przypadku, gdy wyekstrahowano zbyt wiele tekstu dla sformułowania użytego do oceny aspektu, lecz dana fraza faktycznie wyraża opinię autora o danym aspekcie, frazę należy zaklasyfikować jako zaakceptowaną.

D.3 Przykładowe adnotacje

Przykład 1

- **para:** (czarny, rozmiar_wyswietlacza)
- **fraza:** czarny ekran
- **kontekst:** Ale gdy się rano obudziłam nadal był >>>czarny ekran<<<<

Opisany przykład powinien zostać przypisany do klasy *odrzucone*. Powodem jest to, że w danym przykładzie nie jest oceniany aspekt wykryty przez regułę i wpisany w parze.

Przykład 2

- **para:** (niska, cena)
- **fraza:** niższą cenę
- **kontekst:** Samsungi mają >>>niższą cenę<<< i moim zdaniem bardziej opłaca się je kupować niż ten złom.

Przykład powinien zostać zaklasyfikowany jako *odrzucony*. Powodem jest to, że opinia, jaką recenzent ma o ocenianym przedmiocie, jest inna niż wynika to z wydźwięku pary. Fraza *niższą cenę* odnosi się nie do przedmiotu recenzji, ale do innych produktów, z którymi przedmiot recenzji jest porównywany.

Przykład 3

- **para:** (obecna, cena)
- **fraza:** obecnej cenie
- **kontekst:** Moim zdaniem w >>>obecnej cenie<<< absolutnie nie opłaca się go kupować.

Przykład należy przypisać do klasy *odrzucone*. Pomimo tego, że z kontekstu jasno wynika, że recenzent ma negatywną opinię o cenie przedmiotu recenzji, nie jest możliwe przypisanie wydźwięku do pary (obecna, cena) bez analizy kontekstu.

Przykład 4

- **para:** (silny, sygnał)
- **fraza:** silny sygnał
- **kontekst:** Poważnym jego atutem jest >>>silny sygnał<<<.

Przykład należy zaklasyfikować jako *zaakceptowany*, ponieważ parze (*silny, sygnał*) można przypisać pozytywny wydźwięk i analiza kontekstu wskazuje, że opinia recenzenta o tym aspekcie recenzowanego przedmiotu też jest pozytywna.

Przykład 5

- **para:** ((wiele, różnoraki), liczba_funkcji)
- **fraza:** wiele różnorakich funkcji
- **kontekst:** Posiada >>>wiele różnorakich funkcji<<<, a przy tym jest bardzo łatwy w obsłudze

Przykład powinien znaleźć się w klasie *zaakceptowany*. Zarówno sama fraza, jak i kontekst, w którym się znalazła, wyraźnie wskazują na to, że użytkownik wprost wyraża swoją opinię o aspekcie *liczba funkcji*.

E Polarność sformułowań służących do wyrażania opinii – wskazówki adnotacji

E.1 Cel procesu adnotacji

Adnotacja ma na celu zbudowanie bazy przykładów, w których do par (*sformułowanie, aspekt*) przypisywane są polarności, tj. jedna z trzech wartości:

- pozytywna – powinna być przypisywana do par, które użyte w zdaniu wyrażają pozytywną opinię o danym aspekcie,
- negatywna – należy oznaczyć nią pary, które wykorzystane w zdaniu wyrażają negatywną opinię o aspekcie z danej pary,
- neutralna – ta wartość powinna być przypisana do par, w których recenzenci wyrażają swoją opinię o danym aspekcie, ale nie można danych par przypisać do żadnej z powyższych wartości.

Tak zbudowana baza przykładów zostanie następnie wykorzystana do nauki i testowania klasyfikatora, który w sposób automatyczny, na podstawie analizy określonych statystyk, będzie przypisywał pary (*sformułowanie, aspekt*) do klas odpowiadających zdefiniowanym wyżej wartościom.

E.2 Dane do adnotacji

Adnotatorom udostępniane są dwa pliki:

- phrases.ods,
- summary.ods.

Oba pliki zapisane są w formacie OASIS OpenDocument (ods). Mogą one być analizowane i edytowane przykładowo za pomocą programu LibreOffice Calc w wersji 4.2.2.1.

Plik summary.ods zawiera pary (*sformułowanie, aspekt*) i ma następującą strukturę:

- (A) sformułowanie – sformułowanie, za pomocą którego wyrażana jest opinia; jest to pierwszy człon pary (*sformułowanie, aspekt*) i składa się z od jednego do trzech wyrazów sprowadzonych do podstawowej formy (np. w przypadku przymiotnika jest to mianownik liczby pojedynczej w stopniu równym),

- (B) aspekt – aspekt, którego dotyczy dana para (drugi człón pary (*sformułowanie, aspekt*),
- (C) adnotacja – w tej kolumnie ma znaleźć się adnotacja, przypisująca parę do jednej z trzech klas.

Plik posortowany jest rosnąco po nazwach aspektów, a w ramach par o tym samym aspekcie rosnąco po sformułowaniach.

Plik `phrases.ods` zawiera szczegółowe informacje dotyczące wystąpień w korpusie każdej z par z pliku `summary.ods`. Znajdują się tam wszystkie wyekstrahowane frazy i to na podstawie tego pliku stworzony został plik `summary.ods` (dla każdego wiersza z pliku `summary.ods` może być wiele wierszy w pliku `phrases.ods`). Kolejne kolumny w tym pliku mają następującą interpretację:

- (A) sformułowanie – kolumna o interpretacji identycznej do kolumny A w pliku `summary.ods`,
- (B) aspekt – kolumna o interpretacji identycznej do kolumny B w pliku `summary.ods`,
- (C) fraza – cała wyekstrahowana fraza, tak jak wystąpiła w tekście,
- (D) kontekst (ang. *snippet*) – fragment recenzji, z której została wyekstrahowana dana fraza, ukazujący szerszy kontekst, w którym została ona użyta przez recenzenta, obejmujący 50 znaków przed rozpoczęciem danej frazy i 50 znaków po jej zakończeniu.

Plik jest posortowany tak samo jak `summary.ods`, tj. rosnąco po aspektach, a następnie rosnąco po sformułowaniach.

E.3 Wskazówki adnotacji

Zadaniem adnotatorów jest analizowanie kolejnych wierszy w pliku `summary.ods` i przypisywanie do nich wydźwięku. Adnotacje nadawane są w postaci cyfr i mają następującą interpretację:

- 1 – wydźwięk pozytywny,
- 2 – wydźwięk negatywny,
- 3 – wydźwięk neutralny,
- 4 – brak wydźwięku (jeśli sformułowanie nie służy do oceny aspektu).

Adnotacje mają być nadawane w pliku `summary.ods`, natomiast plik `phrases.ods` zawiera dane pomocnicze, służące do określenia polarności w przypadkach, gdy wydzźwięk danej pary (*sformułowanie*, *aspekt*) nie jest dla adnotatora oczywisty.

W trakcie pracy adnotatorzy mają postępować zgodnie z procedurą zdefiniowaną poniżej.

1. Adnotator analizuje wiersz w pliku `summary.ods` oraz identyfikuje te wiersze w pliku `phrases.ods`, które odpowiadają wystąpieniom danej pary. Adnotator przypisuje do pary polarność *a priori* na podstawie wydzźwięku, który zgodnie z jego intuicją oraz znajomością języka i domeny miałyby proste zdania, w których dana para zostałaby wykorzystana. Przykładowo, adnotator może skonstruować zdanie „*aspekt* jest *sformułowanie*” (np. dla pary (światny, bateria) byłoby to „bateria jest światna”) i ustalić polarność *a priori* na podstawie wydzźwięku, jaki jego zdaniem należy przypisać do tego właśnie zdania.
2. Dla zidentyfikowanych w pliku `phrases.ods` wierszy (lub ich 10-elementowej próbki w przypadku par, które zostały wyekstrahowane więcej niż dziesięć razy) adnotator analizuje kolumny C oraz D (*fraza* i *kontekst*) i ustala, jakie opinie wyrażali recenzenci za pomocą danego sformułowania o danym aspekcie. Może się okazać, że część z wierszy została błędnie wyekstrahowana i nie jest wyrażana w nich opinia o danym aspekcie. Takie wiersze ignorujemy i nie bierzemy ich dalej pod uwagę przy ustalaniu wydzźwięku.
 - Jeśli w każdym z analizowanych wierszy wydzźwięk był taki sam i był zgodny z polarnością *a priori*, adnotator przypisuje tę polarność do pary w pliku `summary.ods` i przechodzi do kolejnej pary.
 - Jeśli nie ma pełnej zgodności wydzźwięków w analizowanych wierszach, przyjmujemy następującą strategię. Obliczamy, jaki procent wszystkich wierszy stanowią te o danym wydzźwięku. Jeśli otrzymana wartość jest równa co najmniej 60% i jest dany wydzźwięk jest zgodny z polarnością *a priori*, przypisujemy daną polarność do pary w pliku `summary.ods` i przechodzimy do kolejnej pary.
 - W przypadku braku możliwości przypisania polarności po zastosowaniu opisanej procedury, adnotator, o ile to możliwe, może rozszerzyć analizę na większą liczbę wierszy z pliku `phrases.csv`, np. na 20 wierszy (jeśli dana para została wyekstrahowana więcej niż 10 razy). Gdy nie ma takiej możliwości, adnotator powinien wykorzystać swoją intuicję i wiedzę o domenie i języku w celu ustalenia, czy nadanie polarności *a priori*

jako adnotacji jest uzasadnione. Jeśli tak, adnotator nadaje adnotację i przechodzi do kolejnej pary.

3. W przypadkach, gdy adnotator nie jest w stanie przypisać polarności zgodnie z opisanymi powyżej regułami, dana para powinna zostać zaadnotowana jako „brak wydźwięku” (za pomocą cyfry 4).

F Analiza wydźwięku – wskazówki adnotacji

F.1 Cel procesu adnotacji

Adnotacja ma na celu przygotowanie kolekcji przykładów, składających się z wycinków tekstu i informacji o tym, jaka opinia o zadanym aspekcie została w tym wycinku wyrażona. Za pomocą tak przygotowanego korpusu będzie możliwe oszacowanie poprawności przeprowadzania analizy wydźwięku z wykorzystaniem leksykonów z polarnościami przypisanymi przez różne analizowane metody.

F.2 Dane do adnotacji

Adnotatorom zostanie udostępniony plik zapisany w formacie OASIS OpenDocument (ods). Może on być analizowany i edytowany przykładowo za pomocą programu LibreOffice Calc w wersji 4.2.2.1. Plik zawiera fragmenty tekstów wyekstrahowanych z recenzji telefonów komórkowych. Do każdego wycinka zdefiniowany jest aspekt (pełna lista aspektów znajduje się w tabeli 1 na stronie 181).

W pliku znajdują się następujące kolumny:

- kontekst – wycinek tekstu do przeanalizowania,
- aspekt – aspekt, opinię o którym w tekście adnotator ma ustalić,
- sformułowanie – wyekstrahowane sformułowanie, za pomocą którego prawdopodobnie recenzent wyrażał swoją opinię o aspekcie
- anno – w tej kolumnie adnotator ma umieścić swoją adnotację.

F.3 Wskazówki adnotacji

Do każdego wycinku tekstu adnotatorzy mają przypisać jedną wartość, która ich zdaniem najlepiej opisuje opinię recenzenta o zadanym aspekcie. Te wartości to:

1. opinia pozytywna,
2. opinia negatywna,
3. opinia neutralna lub brak opinii.

Co ważne, adnotacja ma być przypisana na podstawie analizy całego wycinka tekstu. W związku z tym, analiza samej pary (*sformułowanie, aspekt*) nie jest wystarczająca, gdyż kontekst, w którym dana para została użyta, może zmieniać jej sens. Poniżej omówiono wybrane przykłady tego, w jaki sposób adnotacje mają być nadawane.

Przykład 1

- fraza: *wbudowany aparat fotograficzny jest słaby, ale dla mnie wystarczy.*
- aspekt: *aparat fotograficzny*

Ten przykład jest o tyle problematyczny, że recenzent przyznaje, że oceniany aspekt należy nisko ocenić, ale dla niego subiektywnie nie jest to wada. Pomimo tego, że dla recenzenta fakt niskiej oceny aspektu nie jest wadą nie zmienia faktu, że recenzent prawdopodobnie oceniłby aspekt w danym produkcie niżej niż ten sam aspekt w większości innych produktów. W związku z tym należy uznać, że opinia recenzenta o aspekcie jest negatywna.

Przykład 2

- fraza: *polecam gorąco, jak za taką cenę jest to rewelacja!*
- aspekt: *cena*

W tym przypadku, pomimo tego, że nie jest możliwe wskazanie żadnego słowa, które bezpośrednio ocenia aspekt cena (wyekstrahowano słowo „taką”, które może mieć różną interpretację w zależności od kontekstu), z kontekstu wynika, że z punktu widzenia postrzegania jakości całego produktu przez recenzenta cena jest dla niego atrakcyjna. W związku z tym opinia recenzenta o aspekcie jest pozytywna.

Przykład 3

- fraza: *jakość dźwięku jest delikatnie gorsza niż w XYZ, ale i tak należy ją ocenić bardzo wysoko*
- aspekt: *jakość dźwięku*

Opinia recenzenta o aspekcie jest pozytywna, pomimo tego, że przyznaje, że są produkty, gdzie dany aspekt może być oceniony jeszcze lepiej. Nie zmienia to jednak tego, jaka adnotacja powinna zostać nadana.

Przykład 4

- fraza: *Posiada bardzo kolorowy, jasny wyświetlacz który dobrze widać nawet w ostrym słońcu*
- aspekt: *rozmiar wyświetlacza*

Zdanie wyraża pozytywną opinię, ale nie dotyczy aspektu, który wyszczególniono w pliku, to jest rozmiaru wyświetlacza. W takim przypadku należy nadać adnotację *opinia neutralna / brak opinii*.

G Przykład wykorzystanej reguły ekstrakcji w formalizmie JAPE

```
MACRO: ASPEKT_RZECZOWNIK_DO_PRZYIMKOW
```

```
(  
(  
{  
Lookup,  
subst.przypadek == "acc"  
}  
|  
{  
Lookup,  
subst.przypadek == "loc"  
}  
)  
)
```

```
Rule:przymiotnik_przyimek_rzeczownik
```

```
Priority:20
```

```
(  
(  
{adj}  
) :opinionWordLabel  
{SpaceToken}  
{prep}  
{SpaceToken}  
(ASPEKT_RZECZOWNIK_DO_PRZYIMKOW) :aspectNameLabel  
) : label  
-->  
:label.opinion_phrase_candidate =  
{opinion_word_base = :opinionWordLabel.adj.base,  
 opinion_word_pos = :opinionWordLabel.adj.full_tag,  
rule="przymiotnik_przyimek_rzeczownik",  
aspect_name = :aspectNameLabel.Lookup.aspectName,  
opinion_word_string = :opinionWordLabel.num.string}
```
